MEDIDA Y PROBABILIDAD

Miguel Ángel García Álvarez

Contenido

Prólogo

El Cálculo Estocástico se encuentra actualmente en el centro de la Teoría de los Procesos Estocásticos, la cual, a su vez, forma parte de la Teoría de la Probabilidad. Esta última se encuentra bastante desarrollada, aunque, en su formulación moderna, aún no cumple 100 años. Tuvo un impulso enorme durante los primeros 33 años del siglo XX al vincularse con la Teoría de la Medida, la cual surgió y se desarrolló ampliamente también durante el mismo periodo. Esta última tuvo su origen en el Cálculo Diferencial e Integral, el cual tiene una historia más antigua.

El origen de la Teoría de la Probabilidad se encuentra en el Cálculo de Probabilidades, el cual surgió con las soluciones que dieron Blaise Pascal, Pierre Fermat y Christiaan Huygens a algunos problemas relacionados con juegos de dados; Pascal y Fermat en el año 1654 ([**33**]) y Huygens en el año 1657 ([**49**]). Algunos años más tarde (1713) se publicó el trabajo de Jacques Bernoulli ([**4**]) en el cual estableció una relación matemática entre la probabilidad de un evento y la frecuencia relativa con la cual ocurre, relación conocida ahora como Teorema de Bernoulli. Fue el primero de los llamados teoremas límite del Cálculo de Probabilidades. El segundo de estos teoremas lo demostró Abraham de Moivre en el año 1733 ([**28**]) y en el estableció lo que años más tarde, en una formulación más general, se llamaría el Teorema Central del Límite. Los teoremas de Bernoulli y de de Moivre marcaron la pauta para el desarrollo del Cálculo de Probabilidades durante un periodo de aproximadamente 200 años. Fue entre finales del siglo XIX y principios del siglo XX cuando los teoremas límite fueron formulados en toda su generalidad. Todos ellos se refieren al comportamiento en el límite de determinadas relaciones que se obtienen de una sucesión de variables aleatorias independientes.

Resulta curioso que haya surgido un Cálculo de Probabilidades en la época en que iba formándose una concepción mecanicista y determinista del mundo, aunque también fue una época en la cual se derrumbaron viejos paradigmas al cuestionarse el pensamiento y las ideas que estaban establecidas como verdades eternas. Era una época revolucionaria en la cual mucha gente decía "no"a lo que se presentaba como la única realidad posible. En 1654 se había iniciado ya la revolución científica del Renacimiento con las publicaciones de algunas de las obras que serían los fundamentos de la ciencia moderna: *Sobre las revoluciones de los cuerpos celestes* (1543), de Nicolás Copérnico, donde, en contraposición a la concepción aritotélica, planteó que es la Tierra la que se mueve alrededor del Sol; *Novum Organum o Indicaciones relativas a la interpretación de la naturaleza* (1620), de Francis Bacon, donde crítica los fundamentos de la filosofía aristotélica; *Discurso del Método, para dirigir bien la razón y buscar la verdad en las ciencias* (1637), de René Descartes, obra en la cual pone en primer plano la duda y la razón y *Discursos y demostraciones matemáticas en torno a dos nuevas ciencias* (1638), de Galileo Galilei, una de las obras en las cuales se funda el método científico y la Física moderna. Todo esto en forma paralela al cuestionamiento del sistema social imperante.

Los filósofos de la antigua Grecia habían tratado el tema del azar. Aristóteles lo trató explicitamente. Consideraba que hay causas accidentales, dentro de las cuales se encuentran tychē ($\tau\acute{\upsilon}\chi\eta$) y autómaton ($\alpha\grave{\upsilon}\tau\acute{o}\mu\alpha\tau o\nu$), las cuales se presentan en cierto tipo de acontecimientos. Tychē puede traducirse como suerte, restringiendo su significado a un encadenamiento de sucesos, relacionados con una persona o grupo de personas, los cuales llevan a una situación no planeada como objetivo por esa persona o grupo. Autómaton podría traducirse como casualidad, entendiendo ésta como una combinación de circunstancias que no se pueden prever y que llevan a un resultado no intencionado. Decía Aristóteles: "... puesto que vemos que unos eventos suceden de la misma manera siempre, otros, en la mayoría de los casos, es claro que en ninguno de ellos se dice que tychē sea su causa, ni que ocurren por tychē, ni en lo que ocurre por necesidad y siempre, ni en lo que sucede la mayoría de las veces. Pero dado que hay eventos que son opuestos a éstos y de los que todos dicen que existen por tychē, es claro que la tychē y el autómaton existen, pues sabemos que tales eventos se dan por tychē y que por tychē son tales." ([**1**], p. 35). En otro texto decía: "es claro que no hay ciencia del accidente. Pues toda ciencia es de lo que o se da siempre o habitualmente. De lo contrario, ¿cómo se podría enseñar a otro? Es preciso, en efecto, que esté definido o por el siempre o el habitalmente."([**2**], p. 102).

Una idea más de Aristóteles refutada, surgió la ciencia del azar. Lo que en una época parecía que no podía ser, fue, cuando el pensamiento se abrió a nuevas posibilidades.

Pero llevó tiempo y esfuerzo de mucha gente para que la ciencia del azar ocupara un lugar dentro de la Matemática. De hecho, a principios del siglo XX, el Cálculo de Probabilidades era considerado como parte de la Física, no de la Matemática. Así lo dejó ver David Hilbert en el Congreso Internacional de Matemáticas de 1900, donde expresó: "Pienso que en cualquier lugar en donde se presenten ideas matemáticas, sea en Filosofía, sea en Geometría, sea en Física, se plantea el problema de la discusión de los principios fundamentales, base de esas ideas, y del establecimiento de un sistema simple y completo de axiomas... Las investigaciones sobre los principios fundamentales de la geometría nos conducen a plantear este problema: Tratar con base en ese modelo las ramas de la Física donde las Matemáticas juegan actualmente un papel preponderante; esas ramas de la ciencia son, antes que cualesquiera otras, el Cálculo de Probabilidades y la Mecánica".([**48**])

Incluso, a pesar del desarrollo del Cálculo de Probabilidades, por lo menos hasta mediados del siglo XIX el azar parecía un concepto que en algún momento perdería importancia pues, para los pensadores de la época, sólo era producto de nuestra ignorancia. Pierre Simon

Laplace formuló esta idea de manera muy clara, en un artículo publicado en el año 1814, al afirmar que "todos los acontecimientos, aun aquellos que por su insignificancia parecen no depender de las grandes leyes de la naturaleza, constituyen una sucesión tan necesaria como las revoluciones del Sol. Ignorando los vínculos que los ligan al sistema entero del universo, se los ha hecho depender de causas finales o del azar, según que ocurrieran y se sucedieran con regularidad o sin orden aparente; pero esas causas imaginarias han retrocedido gradualmente con los límites de nuestros conocimientos y desaparecen por completo frente a la sana filosofía que no ve en ellas más que la expresión de nuestra ignorancia respecto de las verdaderas causas ... una inteligencia que en un determinado instante pudiera conocer todas las fuerzas que impulsan la naturaleza y la respectiva posición de los seres que la componen y que, además tuviera la suficiente amplitud para someter esos datos al análisis, incluiría en una sola fórmula los movimientos de los mayores cuerpos del universo y del más ligero átomo; nada le sería incierto y tanto el pasado como el futuro estarían en su presencia." ([**56**], p. 2-3)

Sin embargo, con los avances científicos de la segunda mitad del siglo XIX, la concepción determinista mecanicista comenzó a resquebrajarse y un nuevo paradigma comenzó a gestarse, el cual se afianzaría plenamente durante el siglo XX. Durante la segunda mitad del siglo XIX surgió la Mecánica Estadística con los trabajos de Krönig , Clausius , Maxwell y Boltzmann, donde el Cálculo de probabilidades se constituyó como la herramienta fundamental para el estudio de sistemas con muchas partículas. Además, fue en ese periodo cuando surgió la teoría de Mendel sobre la herencia y la teoría de Darwin sobre la evolución de las especies, la primera fundada en un modelo probabilístico y la segunda planteando que el surgimiento de nuevas especies se realiza al azar. Más aún, los estudios de datos, los cuales eran tratados con métodos probabilísticos, crecieron a un ritmo acelerado con los trabajos de Bienaymé, Quetelet y Galton, entre otros. La nueva irrupción del azar en el pensamiento científico cambió el paradigma; de ser pensado únicamente como un producto de nuestra ignorancia, pasó a conceptualizarse como algo objetivo. En 1896, Poincaré expresó claramente este cambio:

"... en la teoría cinética de los gases, se encuentran las conocidas leyes de Mariotte y de Gay-Lussac, gracias a la hipótesis de que las velocidades de las moléculas gaseosas varían irregularmente, es decir, al azar. Las leyes observables serían mucho menos simples, dirían los físicos, si las velocidades estuvieran arregladas por alguna ley elemental simple, si las moléculas estuvieran, como se dice, organizadas, si obedecieran a alguna disciplina. Es gracias al azar, es decir, gracias a nuestra ignorancia, que podemos concluir; y entonces, si la palabra azar es simplemente un sinónimo de ignorancia, ¿qué querría decir eso? ¿Se traduciría entonces como sigue? Me pide usted que le prediga los fenómenos que van a producirse. Si, por desgracia, conociera las leyes de esos fenómenos, podría lograrlo únicamente mediante cálculos inextricables y debería renunciar a responderle; pero, como tengo la suerte de ignorarlas, le voy a responder en seguida. Y, lo más extraordinario, es que mi respuesta será correcta. Se requiere entonces que el azar sea más que el nombre que le damos a nuestra ignorancia." ([**75**], p. 2-3)

A finales del siglo XIX el Cáculo de Probabilidades y el Cálculo Diferencial e Integral éran dos áreas del conocimiento un tanto independientes. La segunda era utilizada por la primera para resolver algunos problemas de probabilidad, así como en otro tipo de problemas se utilizaba el Cálculo Combinatorio. Pero las investigaciones que se llevaban a cabo en uno y otro lado parecían muy alejadas unas de otras; por el lado del Cálculo de Probabilidades se trabajaba principalmente en el estudio de las sucesiones de variables aleatorias independientes (teoremas límite), concepto que ni siquiera existía en el Cálculo Diferencial e Integral. Por el lado de este último las investigaciones estaban orientadas a resolver el problema que había planteado Bernhard Riemann en un artículo publicado en el año 1854 ([**77**]), el cual consistía en caracterizar a las funciones que son integrables de acuerdo con la definición que él mismo dio en su artículo; problema que, a su vez, tenía como objetivo el determinar las condiciones para que una función se pueda expresar como una serie trigonométrica; algo muy alejado del Cálculo de Probabilidades. Nadie imaginaba que pudiera haber un vículo estrecho entra ambas áreas.

Pero las cosas fueron cambiando y tomaron un rumbo inesperado. Las investigaciones alrededor del problema planteado por Riemann condujeron a caracterizar a las funciones integrables en términos de un concepto que surgió en el transcurso de esas investigaciones, el de conjunto de contenido cero: una función acotada $f : [a, b] \to \mathbb{R}$ es Riemann integrable si y sólo si, dada cualquier $\varepsilon > 0$, el conjunto de puntos en los cuales la oscilación de la función es mayor que ε tiene contenido cero. Para aclarar lo que dice este resultado, demos las correspondientes definiciones: Si $f : [a, b] \to \mathbb{R}$ es una función acotada, $x \in (a, b)$ y $[a_1, b_1], [a_2, b_2], \ldots$ es una sucesión de intervalos cerrados encajados que contienen a x como punto interior y tales que $\bigcap_{n=1}^{\infty} [a_n, b_n] = \{x\}$, el límite $\text{lím}_{n\to\infty} [\sup \{f(x) : x \in [a_n, b_n]\} - \text{ínf} \{f(x) : x \in [a_n, b_n]\}]$ existe y es independiente de la sucesión particular de intervalos encajados con las propiedades dadas antes; a ese límite se le llama la oscilación de la función f en el punto x. Si A es un subconjunto de $\mathbb{R}$, se dice que A tiene contenido cero si, dada cualquier $\varepsilon > 0$, existe un número finito de intervalos abiertos cuya unión contiene al conjunto A y tales que la suma de sus longitudes es menor que ε. Una vez surgido este concepto y dada su importancia en la teoría de integración se desarrolló una Teoría del Contenido, mediante la cual se extiende el concepto de longitud a una familia bastante grande de subconjuntos de $\mathbb{R}$. Esta caracterización jugó un papel muy importante en el desarrollo posterior de la teoría de integración; además, echó por tierra algunas ideas que se tenían y significó un giro fundamental en las investigaciones que se venían realizando. Antes de esta caracterización de la integrabilidad en términos del concepto de contenido cero se pensaba que el que una función fuera integrable tenía que ver con el tamaño topológico del conjunto de sus discontinuidades, o, planteándolo de una manera más general, se pensaba que tenía que ver con propiedades topológicas del conjunto de sus discontinuidades. Especificamente, la caracterización de las funciones integrables intentó hacerse con base en dos conceptos topológicos: el de conjunto denso en ninguna parte y el de conjunto de primera especie. Recordemos que se dice que un conjunto A de números reales es denso en ninguna parte si dado cualquier intervalo abierto no vacío I, existe un intervalo abierto no vacío J contenido en I tal que $J \subset A^c$. Esto es equivalente a decir que la cerradura de A no tiene puntos interiores. Por otra parte, si denotamos por $A^{(1)}$ al conjunto de puntos de acumulación de A, por $A^{(2)}$ al conjunto de puntos de acumulación de $A^{(1)}$, etcétera. Al conjunto $A^{(n)}$ se le llama el enésimo conjunto derivado de A. Se dice que A es de primera especie si $A^{(n)}$ es finito para alguna n. Hacia 1873 era ya bien conocido que un conjunto acotado de primera especie es denso en ninguna parte, sin embargo, se pensaba que los conjuntos de primera especie agotaban las posibilidades de los conjuntos densos en ninguna parte, es decir se pensaba que un conjunto es denso en ninguna parte si y sólo si es de primera especie. La confusión terminó cuando se inventaron métodos para construir conjuntos densos en ninguna parte, con lo cual se encontraron ejemplos de conjuntos densos en ninguna parte que no son de primera especie. El concepto de contenido cero no es topológico, se deriva del concepto de longitud; sin embargo, se tiene la siguiente relación: todo conjunto acotado de primera especie tiene contenido cero y a su vez todo conjunto acotado de contenido cero es denso en ninguna parte.

La Teoría del Contenido fue desarrollada principalmente por Camille Jordan alrededor del año 1890. De lo que se trataba era de definir la longitud de cualquier subconjunto de $\mathbb{R}$ y el área de cualquier subconjunto de $\mathbb{R}^2$. Jordan dio un método para lograr esto definiendo el contenido interior de un conjunto (acercándonos por adentro a lo que podría entenderse por área del conjunto, mediante uniones finitas de intervalos, en el caso de $\mathbb{R}$, y por uniones finitas de rectángulos, en el caso de $\mathbb{R}^2$) y el contenido exterior de un conjunto (acercándonos al área que se quiere definir mediante uniones finitas de intervalos que cubran al conjunto, en el caso de $\mathbb{R}$, y mediante uniones finitas de rectángulos que cubran al conjunto, en el caso de $\mathbb{R}^2$). El contenido (longitud o área) de un conjunto estará bien definida cuando su contenido interior coincida con su contenido exterior. La familia de conjuntos cuyo contenido está bien definido no incluye a todos los subconjuntos de $\mathbb{R}$ o de $\mathbb{R}^2$, según sea el caso, pero es bastante grande, tiene la misma cardinalidad que la familia de todos los subconjuntos de $\mathbb{R}$ o de $\mathbb{R}^2$, según sea el caso. La función Contenido tiene la propiedad siguiente: Si se tiene una familia finita de conjuntos ajenos por parejas y cuyo contenido está bien definido, entonces el contenido de la unión también está bien definido y es igual a la suma de los contenidos de los conjuntos de la familia. A esta propiedad la llamaremos la propiedad de la aditividad finita.

Años más tarde, buscando resolver un problema que no tenía nada que ver con la Teoría de Integración, ni mucho menos con el Cálculo de Probabilidades, Émile Borel definió el concepto de medida cero: Si A es un subconjunto de $\mathbb{R}$, se dice que A tiene medida cero si, dada cualquier $\varepsilon > 0$, existe un número finito o infinito numerable de intervalos abiertos cuya unión contiene al conjunto A y tales que la suma de sus longitudes es menor que ε. Una vez definido este concepto, Borel se planteó el problema de desarrollar una teoría de la medida mediante la cual se pudiera extender el concepto de longitud a una familia más grande de subconjuntos de $\mathbb{R}$ que la familia a la que se llegaba con el concepto de contenido. Borel no logró su objetivo, pero sí lo hizo Henri Lebesgue, quien desarrolló lo que ahora se llama la Teoría de la Medida de Lebesgue, mediante la cual se asigna una medida (longitud) a cada elemento de una familia de subconjuntos de los números reales, la cual es más grande que la familia a la que se llega con el concepto de contenido. Esta medida tiene la propiedad siguiente: Si $(A_n)_{n\in\mathbb{N}}$ es una sucesión de subconjuntos de $\mathbb{R}$ a cada uno de los cuales se les puede asignar una medida y esos conjuntos son ajenos por parejas, entonces la unión de ellos también tiene asignada una medida, la cual es igual a la suma de las medidas de cada uno de los conjuntos de la sucesión. A esta propiedad la llamaremos la propiedad de la aditividad numerable o σ-aditividad.

¿Qué vínculo podía tener lo anterior con el Cálculo de Probabilidades?

Por el lado del Cálculo de Probabilidades, una de las propiedades básicas con la que se contaba para calcular probabilidades es la que formula Henri Poincaré en un libro publicado en el año 1896: "cuando un evento puede producirse de dos maneras diferentes, de tal forma que esas dos maneras no puedan ocurrir simultáneamente, la probabilidad de ocurrencia de este evento es igual a la suma de la probabilidad de que se produzca de la primer manera y de la probabilidad de que se produzca de la segunda manera." De aquí se sigue inmediatamente que si un evento puede producirse de n maneras diferentes, de tal forma que cualesquiera dos de esas maneras no puedan ocurrir simultáneamente, la probabilidad de ocurrencia de este evento es igual a la suma $p_1 + p_2 + \cdots + p_n$, donde, para $k \in \{1, 2, \ldots, n\}$, p_k es la probabilidad de que se produzca de la k-ésima manera. ¡Es la aditividad finita! Se van acercando las ideas.

¿Qué se podía decir (estamos ubicándonos a finales del siglo XIX) si un evento puede producirse de una infinidad numerable de maneras diferentes, de tal forma que cualesquiera dos de esas maneras no puedan ocurrir simultáneamente? Hacía muchos años que Jacques Bernoulli había dado una respuesta a esta pregunta, en su libro publicado en el año 1713. No lo hizo para el problema general que estamos planteando, sino para un problema particular donde se presentaba esta situación. La respuesta, para el problema particular que planteó Bernoulli, es simple: la probabilidad de ocurrencia de tal evento es la serie $\sum_{k=1}^{\infty} p_k$, donde para $k \in \mathbb{N}$, p_k es la probabilidad de que el evento se produzca de la k-ésima manera. ¡Es la σ-aditividad! Estamos más cerca. Sin embargo, aún muy lejos ya que una propiedad que se presenta en un caso particular no puede ser trasladada mecanicamente a una situación general. Todavía había un camino por recorrer. Además, lo anterior parece simplemente una analogía; en el caso del contenido de Jordan o de la medida de Lebesgue lo que tenemos son subconjuntos de $\mathbb{R}$ (o de $\mathbb{R}^2$), mientras que en el caso de la probabilidad lo que tenemos son eventos. Tienen propiedades similares, pero estamos tratando con dos tipos de objetos.

Sin embargo la historia no finalizó con lo que hizo Lebesgue ni el Cálculo de Probabilidades se estancó en el estado en que se encontraba a finales del siglo XIX. Lebesgue continuó con sus investigaciones y se le fue uniendo más gente. Después de que Lebesgue desarrolló su teoría de integración en $\mathbb{R}$, se extendió al caso de $\mathbb{R}^n$ sin mucha dificultad. ¿Y si tuviéramos un espacio de dimensión infinita? Parece dificil y ¿nos serviría para algo?, bueno, tal vez no se trate de que sirva para algo. Pero regresemos a lo que ocurrió... Bueno, mejor dejemos la historia aquí por el momento, la continuaremos más adelante. Veamos el final, bueno, no el final final, sólo el final de esta parte. Hacia 1930 se tenía ya desarrollada una Teoría General de la Medida, la cual incluye la definición y el estudio de medidas definidas sobre una familia de subconjuntos de un conjunto cualquiera (sí, cualquiera) y se contaba con un método para construir esas medidas. La propiedad básica de cualquier medida es la σ-aditividad.

¿Y qué pasó por el lado del Cálculo de Probabilidades? Antes de responder esta pregunta, adelantémonos un poco. Mediante un razonamiento lógico, podríamos acercar más el Cálculo de Probabilidades a la Teoría de la Medida, salvando uno de los obstáculos planteados antes. Un problema de probabilidad lo podemos plantear en términos de lo que se denomina un experimento aleatorio, el cual se define como un proceso cualquiera que conduzca a un resultado, pero con la característica de que ese resultado no está únicamente determinado, puede ser uno cualquiera de un conjunto de posibles resultados. ¿Conjunto?, Sí, conjunto. Por fin los conjuntos en el Cálculo de Probabilidades. Al conjunto de posibles resultados de un experimento aleatorio se le denomina espacio muestral y se le suele denotar por la letra Ω. Ahora bien, ¿qué es un evento?, podríamos definirlo como una proposición (en el sentido de la Lógica, es decir una aseveración que se hace la cual únicamente puede ser verdadera o falsa) relativa al resultado del experimento aleatorio que estemos considerando. Si realizamos el experimento podremos decir si esa proposición es verdadera o falsa (para esa realización); utilizando la terminología del Cálculo de Probabilidades, cuando la proposición resulte verdadera diremos que el evento ocurre y cuando es falsa diremos que no ocurre. Dado un evento cualquiera, hay un conjunto de posibles resultados del experimento para los cuales la proposición que define al evento resulta verdadera y para cualquier posible resultado que no esté en ese conjunto la proposición resulta falsa. De esta forma podemos identificar al evento en consideración con ese conjunto, el cual es un subconjunto del espacio muestral. Ah!, entonces la probabilidad es una función que está definida sobre una familia de subconjuntos de un conjunto; como una medida. Salvado un obstáculo, ya no estamos tratando con dos tipos de objetos. Queda entonces un único punto que salvar para poder identificar una función de probabilidad con una medida. ¿Es σ-aditiva cualquier función de probabilidad?

Volvamos a la historia. Casi inmediatamente después de publicado el trabajo de Lebesgue, la naciente Teoría de la Medida comenzó a utilizarse en algunos problemas de probabilidad, por ejemplo para calcular un tipo de probabilidades llamadas geométricas, las cuales consisten en considerar la elección al azar de un punto en una determinada región del plano y calcular la probabilidad de que el punto seleccionado pertenezca a un subconjunto dado de esa región. La probabilidad buscada se calculaba simplemente dividiendo el área del subconjunto dado entre el área de la región (la cual obviamente tendría que ser positiva). Con la teoría de Lebesgue ese problema podía ser resuelto para una familia más grande de subconjuntos de la región donde se selecciona el punto. En ese caso la función de probabilidad resulta ser σ aditiva ya que está definida mediante la medida de Lebesgue en el plano. Pero, se trataba únicamente de un tipo de problemas de probabilidad.

En el año 1909 Borel publicó un artículo donde trató un tipo de problemas que denominó de probabilidades numerables. Para no tener que dar nuevas definiciones, podemos plantear un ejemplo del tipo de problemas que Borel estudió en ese artículo. Supongamos que se lanza un dado una infinidad de veces, ¿cuál es la probabilidad de que el número 6 se obtenga una infinidad de veces? Si se asume que la función de probabilidad es σ-aditiva, la respuesta es que esa probabilidad es igual a 1. Borel llegó a este resultado pero planteándolo en otros términos, sin asumir que la función de probabilidad es σ-aditiva; escribió Borel: "es claro que no se puede buscar aquí la probabilidad de que el caso favorable se produzca una infinidad de veces en n ensayos y enseguida hacer crecer n indefinidamente; por lo tanto se razonará como sigue: eligiendo un número fijo m, se buscará la probabilidad de que el caso favorable se produzca más de m veces en n ensayos y se calculará el límite hacia el cual tiende esta probabilidad cuando n aumenta indefinidamente; omito aquí el sencillo cálculo, cuyo resultado es el siguiente: este límite es la unidad cualquiera que sea el número fijo m; eso significa que se puede apostar con ventaja una cantidad tan grande como se quiera contra 1 franco a que el número de casos favorables será superior a un número fijo dado cualquiera m; es precisamente la significación de este enunciado: la probabilidad $P(A_\infty)$ es igual a uno" (en nuestro ejemplo A_∞ es la obtención del número 6 una infinidad de veces). Aunque pareciera que sí, Borel no afirmó que $P(A_\infty) = 1$, o, si se quiere, lo afirmó pero solamente como un enunciado con una interpretación particular. Y había una razón para que Borel no escribiera $P(A_\infty) = 1$ sin necesidad de dar una determinada interpretación. La razón era que Borel había mostrado que una función de probabilidad no siempre es σ-aditiva. Esto lo hizo dando el siguiente ejemplo: "Supongamos, por ejemplo, que existe una manera de elegir de entre la colección infinita de números enteros, uno de ellos al azar, de manera que cada uno de ellos tenga la misma probabilidad, esta probabilidad deberá entonces ser nula, pero su suma debe ser igual a 1". ¡Ups!, efectivamente no hay σ-aditividad. ¡Qué complicación!, tan bien que íbamos. Alguien podría argumentar que no hay ningún problema pues el experimento aleatorio que planteó Borel es irrealizable, ¿cómo elegir al azar un número entero? La aceptación de ese argumento nos metería en problemas ya que en el Cálculo de Probabilidades nos encontramos con muchos experimentos de ese tipo; son experimentos pensados. Serían serios los problemas pues un experimento de importancia básica en el Cálculo de Probabilidades consiste en la elección al azar de un número real en el intervalo $[0, 1]$, imposible de realizar, pero no de modelar matemáticamente, con una función de probabilidad σ-aditiva. Cabe mencionar aquí que el problema de la imposibilidad de efectivamente realizar un experimento aleatorio no crea ningún problema en la formulación moderna de la Teoría de la Probabilidad ya que esta formulación consiste en definir y estudiar un sistema matemático formal, con reglas muy precisas que no tienen nada que ver con la realización de experimentos; dentro de ese sistema, todo es Matemática. La utilización de ese cuerpo teórico formal para modelar fenómenos que se presentan en la realidad es un problema aparte.

Pero sigamos con la historia. Se fueron planteando problemas de probabilidad cada vez más complejos, ya no únicamente con sucesiones de variables aleatorias independientes. Un problema de gran importancia que se resolvió fue el de construir un modelo matemático (probabilístico) para el llamado movimiento browniano, el cual consiste en el movimiento de un grano de polen que se coloca sobre agua. Las posibles trayectorias que sigue el grano de polen sobre el agua son funciones continuas definidas en el intervalo de tiempo en que se observa el movimiento, el cual podríamos asumir que es el intervalo $[0, \infty)$ y podríamos imaginar que el recipiente de agua es infinito, de manera que cada posible trayectoria que sigue el grano de polen es una función continua $f : [0, \infty) \to \mathbb{R}^2$; así que el conjunto de posibles trayectorias es un espacio vectorial de dimensión dimensión. Fue Norbert Wiener quien, en el año 1923, construyó ese modelo. Para ello utilizó los resultados que había obtenido John Daniell, quien entre los años 1918 y 1920, desarrolló una teoría general de integración la cual permitía definir la integral para funciones que dependen de una infinidad de variables. El método de Daniell conduce a una definición de integral la cual tiene las mismas propiedades que la integral que definió Lebesgue, en particular en lo que se refiere a poder integrar, bajo determinadas condiciones, el límite de una sucesión de funciones como el límite de las integrales de las funciones de la sucesión, propiedad que equivale a la σ-aditividad de una medida. Teníamos nuevamente la σ-aditividad de la función de probabilidad, pero todavía para un caso particular, aunque esta vez de mucha importancia.

El ejemplo de Borel, de una función de probabilidad que no es σ-aditiva, medio se olvidó, tal vez por todos los resultados que se iban obteniendo en la Teoría de la Medida, dándole así mucha fuerza. En el año 1925 se publicó un libro de Paul Lévy titulado *Calcul des Probabilités*, donde asume que toda función de probabilidad es σ-aditiva. Sin embargo aún quedaban algunos obstáculos que salvar para una aceptación general de esta propiedad. El problema central era el siguiente: A una variable aleatoria se le asocia una función no decreciente denominada su función de distribución y, desde el año 1913, August Radon había demostrado que, a partir de una función de ese tipo, se puede construir una medida utilizando el método de Lebesgue; a una familia finita formada por n variables aleatorias se le asocia también una función de n variables denominada su función de distribución conjunta y, utilizando el método que Constantin Carathéodory había publicado en el año 1914, también, a partir de una función de distribución conjunta, se podía construir una medida. El problema que quedaba por resolver era como construir una medida asociada a una infinidad, numerable o no numerable, de variables aleatorias. Este problema lo resolvió, utilizando el método de Carathéodory, Andrei Nikolayevich Kolmogorov en el año 1933. De esta manera la fusión del Cálculo de Probabilidades con la Teoría de la Medida quedaba consumada. El sistema matemático formal que surgió con las investigaciones realizadas entre los años 1900 y 1933, las cuales culminaron con la publicación del trabajo de Kolmogorov, es lo que propiamente podemos llamar ahora la Teoría de la Probabilidad.

¿Y qué hacer con el ejemplo de Borel? Hubo una polémica al respecto que se dio mediante la publicación de varios artículos. La conclusión para quienes optaron por aceptar la σ-aditividad como propiedad de cualquier función de probabilidad, fue simplemente que ese ejemplo queda fuera del cuerpo teórico que se desarrolló.

Además de lo interesante y maravilloso de esta historia, tenemos aquí un ejemplo de entrelazamiento, y en algún momento prácticamente una fusión, entre dos áreas del conocimiento que parecían independientes una de la otra. Una inventada para tratar con sistemas deterministas, como el movimiento de los cuerpos celestes; otra inventada para tratar con procesos que evolucionan al azar.

Analizando la historia, tanto de la Teoría de Integración como del Cálculo de Probabilidades, hay algunos aspectos que merecen ser resaltados:

1. El desarrollo se va dando gracias a que cada uno de los que fueron contribuyendo tuvieron una mirada crítica hacía lo que estaba ya hecho. Algunos ejemplos: Cuando Fourier afirmó

que toda función se puede expresar como una serie trigonométrica, algunos tomaron el resultado cautelosamente y se pusieron a investigar sobre el tema, Cauchy se dio cuenta que una de las cosas que es necesario definir con precisión es el concepto de integral, pero lo hizo únicamente para las funciones continuas; después, él mismo y otros que le siguieron buscaron definir la integral para el caso de funciones discontinuas; la integral se definía dependiendo del tipo de discontinuidades. Más adelante, Riemann, con una posición crítica hacia lo que se venía haciendo, cambió el enfoque al definir la integral de una única manera, independientemente de como sea la función; se abocó entonces al problema de caracterizar a las funciones que son integrables. Borel introdujo el concepto de medida cero e intentó desarrollar una Teoría de la Medida, sin embargo, su planteamiento fue limitado. Lebesgue, mirando el trabajo de Borel criticamente, planteó otro camino que resultó mucho más fructífero. Vemos entonces que es un pensamiento crítico el que va produciendo el desarrollo.

2. En el recuento histórico de los párrafos anteriores quedan muchísimos nombres sin mencionar. Tanto del lado del Cálculo de Probabilidades como del Análisis Matemático, los avances que se hicieron fueron producto del trabajo de mucha gente, lo cual muestra que la matemática es un producto social y que se desarrolla en zigzag, buscando a veces en una dirección sin encontrar resultados suficientemente satisfactorios, dejando preguntas abiertas que en un momento dado no pudieron responderse y que la tenacidad de quienes continuaron permitieron irlas respondiendo, no necesariamente en la forma en que fueron inicialmente formuladas. Se trata de un trabajo colectivo que va creando una historia de suspenso, porque en cada momento no se sabe cómo se llegará a la solución de un determinado problema ni qué nuevos elementos se irán introduciendo para abordarlo. No se sabe tampoco si surgirá una nueva teoría, que tal vez lo englobe o lo descarte; todo esto en un proceso que no tiene fin, porque nunca estará todo dicho, aún cuando la historia pudiera continuar indefinidamente.

3. La formulación que se dio a la Teoría de la Probabilidad es axiomática, es decir, se definió un sistema formal de axiomas, con los cuales se ha desarrollado un cuerpo teórico puramente matemático, independiente de los procesos o fenómenos reales que posibilitaron su surgimiento. ¿Independiente realmente? La respuesta es dialéctica, sí y no.

Sí, porque efectivamente el cuerpo teórico que se va creando tiene su propia dinámica interna; es válido, por ejemplo, inventar definiciones con los elementos de ese cuerpo teórico y construir una teoría con resultados relativos a lo definido, la cual quedaría dentro del cuerpo teórico general; y esto aún cuando lo definido o los resultados que se demuestren no tengan vínculo alguno con algún fenómeno o proceso real.

No, porque lo que ocurre en el desarrollo histórico es que, en general, son los fenómenos reales los que motivan el interés en estudiar un determinado problema. Son los fenómenos reales los que direccionan el desarrollo teórico. A su vez, el cuerpo teórico que se forma a partir de los axiomas es utilizado para resolver algún nuevo problema real. En un contexto más general, el conocimiento va formando un entramado complejo con vínculos entre diferentes áreas, incluyendo lo social. Una investigación científica, de cualquier área, tiene la posibilidad, por su naturaleza de producción, de tener efectos en otras áreas del saber, dentro de cualquiera de sus ramas, las cuales, todas juntas, conforman el entramado simbólico humano.

Por otra parte, como lo mencionamos antes, el cuerpo teórico que conforma la Teoría de la Probabilidad es puramente matemático, basado en un sistema de axiomas, el cual incluye la propiedad de σ-aditividad de la función de probabilidad, de manera que viéndolo retrospectivamente, la pregunta ¿es σ-aditiva cualquier función de probabilidad?, planteada en uno de los párrafos anteriores, está mal formulada, o, si se quiere, tiene una respuesta trivial: cualquier función de probabilidad es σ-aditiva si el sistema de axiomas incluye esa propiedad; de otra forma, no necesariamente se tiene la σ-aditividad. En otras palabras, podríamos definir un sistema axiomático en el cual una función de probabilidad tuviera como propiedad necesaria el ser finitamente aditiva; pudiendo ser σ-aditiva o no serlo. El cuerpo teórico que se desarrollara a partir de ese sistema axiomático conformaría una teoría matemática más, con el mismo status que cualquier otra. El incluir a la σ-aditividad como propiedad de cualquier función de probabilidad tiene la justificación que se fue dando históricamente, pero finalmente es, en sí, arbitraria.

Tal vez alguien podría preguntarse, bueno, pero en un fenómeno aleatorio real, ¿la función de probabilidad es o no es σ-aditiva? Nuestro punto de vista es que, parafraseando a Jacques Lacan, un fenómeno aleatorio real no habla, así que no cuenta con un sistema simbólico. La función de probabilidad es una invención humana, la cual forma parte de todo el simbolismo con el que los humanos tratamos de entender a la naturaleza; pero ese simbolismo es únicamente eso, un simbolismo que no forma parte de la naturaleza en sí misma.

4. Hay un aspecto que no desarrollamos en este libro; es el que se refiere al vínculo que hay entre el avance científico y las necesidades del sistema dominante. En ocasiones suele pensarse que el quehacer de un matemático es un acto solipsista, signado de la más absoluta neutralidad, dejándose artificiosamente de lado que dicho acto es un acto humano y como tal se encuentra inmerso en un marco conceptual, y atravesado por la cultura y la ideología. También se cae frecuentemente en el reduccionismo de pensar que es un acto que tiene garantía de ser "sin consecuencias", haciendo entonces a un lado la importancia vital del desarrollo de las ciencias para bien o para mal de la humanidad. Siempre están presentes el mundo real, los problemas sociales y los intereses de determinados grupos, que parecen estar por fuera del quehacer científico, pero que en realidad no lo están totalmente. Hay una responsabilidad del hombre o mujer de ciencia, incluido el matemático o la matemática, de reflexionar sobre lo que atañe a la investigación científica, abriendo canales de comunicación con todo el abanico del desarrollo científico y social, en pos de alcanzar una visión más completa, insertos en un marco histórico específico. Es incluso así como sería más fructífero ese pensamiento crítico que ha sido desde siempre el impulsor de los pasos cruciales del desarrollo a lo largo de la historia de la humanidad. Ese maravilloso pensamiento crítico que pone sobre el tapete los saberes de los que nos precedieron para que pueda darse lugar a los grandes hitos significativos del salto cualitativo del saber.

Volviendo nuevamente a la parte histórica de los temas que se tratan en este libro y recapitulando lo mencionado con anterioridad, durante los primeros 33 años del siglo XX el Cálculo de Probabilidades se desarrolló enormemente; puede decirse que durante esos años culminó un periodo de desarrollo y comenzó uno nuevo. Paralelamente a la conclusión del estudio de las sucesiones de variables aleatorias independientes se comenzaron a formular problemas que involucraban una infinidad de variables aleatorias las cuales dependían unas de otras de diferentes maneras, iniciándose así el estudio de los procesos estocásticos. Este estudio se hizo ya sobre nuevas bases, con una formulación de la Teoría de la Probabilidad basada en la Teoría General de la Medida.

En este libro se expone la formulación moderna de la Teoría General de la Medida y de la Teoría de Integración con respecto a una medida, para pasar después a la formulación de la Teoría de la Probabilidad basándonos en la Teoría de la Medida. Está pensado como el primero de dos libros, en el segundo de los cuales el objetivo es formular las bases del Cálculo Estocástico mediante la definición de lo que se denomina una integral estocástica y el estudio de sus propiedades.

La intención es lograr una exposición de los diferentes temas de tal manera que el libro sea accesible a estudiantes de licenciatura, sin dejar de lado el rigor matemático que se requiere en cada uno de ellos.

El libro consta de 15 capítulos.

En el primero se expone de manera detallada la historia del desarrollo de la Teoría de Integración hasta llegar a la formulación de Lebesgue.

En el segundo se hace una exposición formal de cómo se construye la medida de Lebesgue en $\mathbb{R}$.

En el tercero se trata el tema de las funciones de variación acotada.

En el cuarto se define y se demuestran las propiedades de la integral de Riemann-Stieltjes, la cual fue de gran importancia para generalizar el trabajo de Lebesgue y llegar a formular una Teoría General de la Medida y de la Integral.

En el quinto de desarrolla formalmente la Teoría General de la Medida.

En el sexto se trata el tema de la construcción de medidas a partir de funciones de variación acotada.

En el séptimo se expone la primera parte de la Teoría General de Integración con respecto a una medida.

En el octavo se expone la segunda parte de la Teoría General de Integración con respecto a una medida.

En el noveno se trata nuevamente el tema de la integral de Stieltjes, pero viéndola como la integral con respecto a la medida generada por una función de variación acotada.

En el décimo se definen y estudian diferentes tipos de convergencia de funciones; tema central ya que, por ejemplo, la integral estocástica se define utilizando la convergencia en el espacio de funciones de cuadrado integrable.

En el décimo primero se estudian los espacios L^p.

En el décimo segundo se expone de manera detallada la historia del desarrollo de la Teoría de Probabilidad hasta llegar a su formulación axiomática.

En el décimo tercero se desarrolla la Teoría de la Probabilidad, considerando a ésta como una medida; se definen los conceptos básicos y se estudian sus propiedades; en particular, se muestra cómo se genera una medida a partir de una función de distribución conjunta.

En el décimo cuarto se estudia el concepto de Esperanza y las leyes de los grandes números.

En el décimo quinto se trata el tema de la construcción de espacios de probabilidad, concluyendo el capítulo con la demostración del teorema de Kolmogorov, el cual permite construir un espacio de probabilidad para una familia cualquiera de variables aleatorias.

En el Apéndice se desarrollan algunos temas de Análisis que se utilizan en el texto: el teorema de Heine-Borel, compacidad en espacios métricos, caracterización de los conjuntos compactos, espacios vectoriales normados, convergencia uniforme y la aproximación de funciones continuas mediante polinomios.

Miguel Ángel García Álvarez
Diciembre de 2018
e-mail: magaz@unam.mx

Notación y terminología

$A \cup B$	Unión de los conjuntos A y B.
$A \cap B$	Intersección de los conjuntos A y B.
$\bigcup_{k=1}^{n} A_k$	Unión de los conjuntos $A_1, \ldots, A_n$.
$\bigcap_{k=1}^{n} A_k$	Intersección de los conjuntos $A_1, \ldots, A_n$.
$(A_n)_{n\in\mathbb{N}}$	Sucesión de los conjuntos $A_1, A_2, \ldots$.
$\bigcup_{k=1}^{\infty} A_k$	Unión de los conjuntos de la sucesión $(A_n)_{n\in\mathbb{N}}$.
$\bigcap_{k=1}^{\infty} A_k$	Intersección de los conjuntos de la sucesión $(A_n)_{n\in\mathbb{N}}$.
A^c	Complemento del conjunto A.
$B - A$	$B \cap A^c$.
$A \times B$	Producto cartesiano de los conjuntos A y B.
$A \subset B$	El conjunto A está contenido en el conjunto B.
$A \supset B$	El conjunto A contiene al conjunto B.
$\emptyset$	Conjunto vacío.
$\mathbb{N}$	Conjunto de los números naturales, sin incluir el cero.
$\mathbb{Z}$	Conjunto de los números enteros.
$\mathbb{Q}$	Conjunto de los números racionales.
$\mathbb{R}$	Conjunto de los números reales.
$\overline{\mathbb{R}}$	$\mathbb{R} \cup \{-\infty, \infty\}$.
$\mathbb{Z}^+$	Conjunto de los números enteros no negativos.
$\mathbb{R}^+$	Conjunto de los números reales no negativos.
$\{n, \ldots, m\}$	Conjunto de números enteros entre n y m inclusive.
$\{n, n+1 \ldots\}$	Conjunto de números enteros mayores o iguales a n.

$\text{mín}(a,b)$	Mínimo entre a y b.
$\text{máx}(a,b)$	Máximo entre a y b.
$x \wedge y$	$\text{mín}(x,y)$.
$x \vee y$	$\text{máx}(x,y)$.
$\text{ínf}\, A$	Ínfimo del conjunto A.
$\sup A$	Supremo del conjunto A.
$(x_n)_{n\in\mathbb{N}}$	Sucesión de los números $x_1, x_2, \ldots$.
$x \rightsquigarrow \alpha$	x tiende al valor α.
$f : A \mapsto B$	función definida sobre el conjunto A, con valores en el conjunto B.
$g \circ f$	Composición de las funciones f y g.
I_A	Función indicadora del conjunto A (igual a 1 sobre A y 0 sobre A^c).
$f\,(x+)$	Límite de la función f cuando la variable tiende a x por la derecha.
$f\,(x-)$	Límite de la función f cuando la variable tiende a x por la izquierda.
$\ln x$	Logaritmo natural de x.
$\|x\|$	Valor absoluto del número real x.
$[[x]]$	Mayor entero menor o igual a x.
x^+	$\text{máx}(x,0)$.
x^-	$\text{máx}(-x,0)$.
(a,b)	Intervalo abierto $\{x\in\mathbb{R} \mid a<x<b\}$.
$[a,b]$	Intervalo cerrado $\{x\in\mathbb{R} \mid a\le x\le b\}$.
$(a,b]$	Intervalo semiabierto $\{x\in\mathbb{R} \mid a<x\le b\}$.
$[a,b)$	Intervalo semiabierto $\{x\in\mathbb{R} \mid a\le x<b\}$.
$\sum_{k=1}^n x_k$	Suma de los números $x_1,\ldots,x_n$.
$\sum_{k=1}^\infty x_k$	$\text{lím}_{n\rightsquigarrow\infty}\sum_{k=1}^n x_k$
$\prod_{k=1}^n x_k$	Producto de los números $x_1,\ldots,x_n$.
$\binom{n}{k}$	Combinaciones de n elementos tomados de k en k ($\frac{n!}{k!(n-k)!}$).

Diremos que una sucesión de números reales $(x_n)_{n\in\mathbb{N}}$ es no decreciente (resp. no creciente) si $x_n \leq x_{n+1}$ (resp, $x_n \geq x_{n+1}$) para cualquier $n \in \mathbb{N}$.

Diremos que una sucesión de números reales $(x_n)_{n\in\mathbb{N}}$ es creciente (resp. decreciente) si $x_n < x_{n+1}$ (resp, $x_n > x_{n+1}$) para cualquier $n \in \mathbb{N}$.

Diremos que una sucesión de funciones $(f_n)_{n\in\mathbb{N}}$, de $\mathbb{R}$ en $\overline{\mathbb{R}}$, es no decreciente (resp. no creciente) si $f(x) \leq f(y)$ (resp, $f(x) \geq f(y)$) para cualquier pareja $x, y \in \mathbb{R}$ tales que $x \leq y$.

Diremos que una sucesión de funciones $(f_n)_{n\in\mathbb{N}}$, de $\mathbb{R}$ en $\overline{\mathbb{R}}$, es creciente (resp. decreciente) si $f(x) < f(y)$ (resp, $f(x) > f(y)$) para cualquier pareja $x, y \in \mathbb{R}$ tales que $x < y$.

Diremos que una sucesión de conjuntos $(A_n)_{n\in\mathbb{N}}$ es creciente (resp. decreciente) si $A_n \subset A_{n+1}$ (resp, $A_n \supset A_{n+1}$) para cualquier $n \in \mathbb{N}$.

CAPÍTULO 1

MEDIDA E INTEGRAL DE LEBESGUE
Desarrollo histórico

El Cálculo Diferencial e Integral fue inventado por Isaac Newton y Gottfried Wilhelm Leibniz a finales del siglo XVII. En su trabajo definieron el concepto de derivada de una función y geometricamente la interpretaban como la pendiente de la tangente a su gráfica. La integral de una función la vieron como la operación inversa de la derivada y geometricamente la interpetaban como el área de la región delimitada, en el intervalo de integración, por la gráfica de la función y el eje horizontal.

A medida que la teoría se fue desarrollando se plantearon problemas cada vez más complejos, los cuales hicieron ver la necesidad de definir los conceptos con mayor precisión y de demostrar resultados con métodos analíticos, en lugar de algunos métodos geométricos que se utilizaban. En particular, la manera en que se trataba con la integral de una función llevó a cuestionamientos acerca de la validez de algunas propiedades que se asumían como válidas. De particular importancia fue el trabajo de Jean-Baptiste Joseph Fourier, publicado en el año 1822 bajo el título *Théorie analytique de la chaleur* ([**35**]). Afirmó ahí que una función arbitraria f, definida y acotada en el intervalo $[-L, L]$, puede representarse mediante una serie trigonométrica de la forma:

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} \left[a_n \cos\frac{n\pi x}{L} + b_n sen\frac{n\pi x}{L}\right].$$

La demostración de Fourier de esta afirmación consiste básicamente en tratar el desarrollo anterior como una ecuación para la cual tendrían que encontrarse los coeficientes a_0, a_n y b_n ($n \in \mathbb{N}$) que la hacen válida. Para esto, integrando entre $-L$ y L ambos lados de la expresión, se obtiene:

$\int_{-L}^{L} f(x)dx = 2a_0L.$

Así que $a_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx.$

Ahora, multiplicando por $\cos\frac{n\pi x}{L}$ ambos lados de la expresión e integrando, se obtiene:

$\int_{-L}^{L} f(x)\cos\frac{n\pi x}{L}dx = a_nL.$

Así que $a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{n\pi x}{L}dx$.

Finalmente, multiplicando por $sen\frac{n\pi x}{L}$ ambos lados de la expresión e integrando, se obtiene:

$\int_{-L}^{L} f(x)sen\frac{n\pi x}{L}dx = b_nL$.

Así que $b_n = \frac{1}{L}\int_{-L}^{L} f(x)sen\frac{n\pi x}{L}dx$.

Fourier argumentaba que las integrales que definen los coeficientes a_0, a_n y b_n están bien definidas pues cada una puede obtenerse mediante el cálculo del área bajo la gráfica de la función correspondiente.

Cabe mencionar que por función arbitraria Fourier no se refería a lo que actualmente se entiende por una función como cualquier correspondencia de un conjunto en otro; sin embargo dentro de las funciones que consideraba incluía no únicamente a las funciones continuas.

Además de la necesidad de clarificar el concepto de función, la demostración de Fourier planteaba los siguientes tres problemas:

(i) Definiendo los coeficientes a_n y b_n como lo hacía Fourier,
¿la serie $\frac{1}{2}a_0 + \sum_{n=1}^{\infty}\left[a_n\cos\frac{n\pi x}{L} + b_n sen\frac{n\pi x}{L}\right]$ converge a $f(x)$?
(ii) ¿Para qué funciones f, las integrales que definen los coeficientes a_n y b_n $(n \in \mathbb{N})$ están definidas?
(iii) ¿Se puede integrar término a término una serie de funciones?

En el año 1823 se publicó el libro de Augustin-Louis Cauchy titulado *Résumé des leçons données à l'École Royale Politechnique sur le calcul infinitésimal* ([**17**]), en el cual trató el problema de la definición de la integral, primero para las funciones continuas y después para funciones con discontinuidades.

En ese trabajo, Cauchy definió el concepto de continuidad básicamente como se conoce actualmente:

Una función definida en un intervalo es continua si para cada x en el intervalo el valor numérico de la diferencia $f(x+\alpha) - f(x)$ decrece indefinidamente con α.

Más adelante formuló la definición analítica de la integral de una función continua, demostrando su existencia:

Sea f una función continua en el intervalo $[a,b]$, entonces las sumas:

$$S = \sum_{k=1}^{n} f(x_{k-1})\,(x_k - x_{k-1}),$$

correspondientes a particiones $P = \{a = x_0 < \cdots < x_n = b\}$ tienden a un límite cuando los elementos $x_k - x_{k-1}$ se hacen infinitamente pequeños; a ese límite se le llama la integral definida de f y se le denota por $\int_a^b f(x)dx$. Se obtiene el mismo límite si se consideran sumas de la forma $S = \sum_{k=1}^{n} f\left[x_{k-1} + \theta_k(x_k - x_{k-1})\right](x_k - x_{k-1})$, donde $\theta_k \in [0,1]$.

Demostró además que si f es una función continua y $F(x) = \int_a^x f(y)dy$, entonces $F'(x_0) = f(x_0)$ para cualquier $x_0 \in (a,b)$.

La integral así definida es conocida actualmente como la integral de Riemann y no como la integral de Cauchy. La razón de esto parece justa pues es el trabajo de Riemann, publicado en el año 1867, el que dió la pauta para desarrollar una Teoría de Integración, la cual a su vez llevaría más tarde a una Teoría del Contenido y finalmente a la moderna Teoría de la Medida.

En trabajos posteriores, Cauchy consideró funciones discontinuas haciendo la aclaración siguiente:

“es necesario observar que las funciones discontinuas introducidas en el Cálculo dejan de ser continuas únicamente para algunos valores de las variables”

Para este tipo de funciones discontinuas extendió el concepto de integral de la siguiente manera:

Si una función es continua en un intervalo $[a,b]$, excepto en un punto c, en una vecindad del cual f puede ser acotada o no, se puede definir la integral de f como el límite:

$$\lim_{h \rightsquigarrow 0} \left[\int_a^{c-h} f(x)dx + \int_{c-h}^{b} f(x)dx \right],$$

cuando éste existe.

En 1829, Johann Peter Gustav Lejeune Dirichlet conjeturó que el método de Cauchy para definir la integral de funciones discontinuas se puede extender a todas las funciones que tengan la siguiente propiedad:

Suponiendo que f está definida en un intervalo $[a,b]$, dadas dos cantidades arbitrarias u y v en ese intervalo, es posible encontrar otras dos cantidades r y s entre u y v tales que la función f es continua en el intervalo $[r,s]$.

Es decir, utilizando la terminología moderna, el conjunto de puntos donde la función es discontinua debe ser denso en ninguna parte. Recordemos que se dice que un conjunto A de números reales es denso en ninguna parte si dado cualquier intervalo abierto no vacío I, existe un intervalo abierto no vacío J contenido en I tal que $J \subset A^c$. Esto es equivalente a decir que la cerradura de A no tiene puntos interiores, o bien que A no es denso en ningún intervalo abierto no vacío.

1.1. La integral de Riemann

Georg Friedrich Bernhard Riemann, en un artículo titulado *Sur la possibilité de représenter une fonction par une série trigonométrique* ([**77**]), el cual fue elaborado en 1854 pero publicado en 1867, cambió el enfoque para atacar el problema de la integración de funciones. Cauchy y quienes le siguieron buscaban extender la definición de la integral a funciones tan discontinuas como fuera posible, pero no partiendo de una definición general sino dando una definición distinta dependiendo del tipo de funciones que se querían integrar. En cambio, Riemann planteó una definición general de la integral para cualquier función y se abocó al problema de caracterizar a las funciones para las cuales esa integral está definida.

Planteaba Riemann:

¿Qué se debe entender por $\int_a^b f(x)dx$*?*

Consideremos una partición $x_0, x_1, \ldots, x_n$ del intervalo $[a, b]$ y definamos $\delta_k = x_k - x_{k-1}$. Si, independientemente de como se elijan las cantidades $\varepsilon_k \in [0, 1]$, las sumas

$$\sum_{k=1}^{n} \delta_k f(x_{k-1} + \varepsilon_k \delta_k)$$

tienden a un límite cuando todas las cantidades δ_k tienden a cero, a ese límite se le llama el valor de la integral definida $\int_a^b f(x)dx$.

Decía Riemann:

> *"Busquemos ahora la extensión y el límite de la definición precedente y hagámonos esta pregunta: ¿En qué casos una función es susceptible de integración?, ¿en qué casos no lo es?"*

Estableció dos criterios, ambos basados en el concepto de oscilación de una función en un intervalo.

DEFINICIÓN 1.1. *Sea* $f : [a, b] \to \mathbb{R}$ *una función acotada. La diferencia:*

$$\sup \{f(x) : x \in [a, b]\} - \text{ínf} \{f(x) : x \in [a, b]\}$$

es llamada la oscilación de f *en el intervalo* $[a, b]$*.*

Criterio R_1

Sea D_k la oscilación de f en el intervalo $[x_{k-1}, x_k]$, entonces:

$$f \text{ es integrable si y sólo si } \quad \text{lím}_{\delta_k \to 0} \sum_k D_k \delta_k = 0.$$

Criterio R_2

Dada $\sigma > 0$ y una partición P, sea $\lambda(P,\sigma)$ la suma de las longitudes de los subintervalos de la partición en los cuales la oscilación de la función es mayor que σ, entonces:

$$f \text{ es integrable si y sólo si } \quad \lim_{\|P\|\to 0} \lambda(P,\sigma) = 0 \text{ para cualquier } \sigma > 0,$$

donde $\|P\|$ es la norma de P.

Este criterio se sigue del criterio R_1 y las siguientes desigualdades:

$\sigma\lambda(P,\sigma) \leq \sum_k D_k\delta_k \leq D\lambda(P,\sigma) + (b-a)\sigma.$

donde D_k es la oscilación de f en el intervalo $[x_{k-1}, x_k]$ y D la oscilación de f en el intervalo $[a,b]$.

El criterio R_2 permitió a Riemann dar un ejemplo de una función integrable con un conjunto denso de discontinuidades:

Sea $M = \left\{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots\right\}$ y, para $x \in [0,\infty)$, definamos:

$$(x) = \begin{cases} 0 & \text{si } x \in M \\ x - m(x) & \text{si } x \notin M \end{cases}$$

donde $m(x)$ es el número entero más cercano a x.

Riemann definió entonces la función $f : [0,1] \to \mathbb{R}$ de la siguiente manera:

$f(x) = \sum_{k=1}^{\infty} \frac{(kx)}{k^2}$

Se puede demostrar que esta función es discontinua en todos todos los puntos x de la forma $x = \frac{m}{2n}$, donde m y n son dos números naturales tales que m y $2n$ son primos entre sí. Además f satisface el criterio R_2 de Riemann y, por lo tanto, es integrable.

Un ejemplo similar, pero más fácil de tratar, es el siguiente:

Consideremos la función $f : [0,1] \to \mathbb{R}$ definida por:

$$f(x) = \begin{cases} \frac{1}{n} & \text{si } x = \frac{m}{n} \text{ con } m,n \in \mathbb{N} \text{ y primos entre sí} \\ 0 & \text{en otro caso} \end{cases}$$

Esta función es continua en los irracionales y discontinua en los racionales. En efecto, la discontinuidad en un número racional x se sigue del hecho de que nos podemos acercar a x mediante números irracionales, en los cuales f toma el valor 0. Para demostrar la contuidad en un número irracional, primero observemos que dada $\varepsilon > 0$, únicamente existe un número finito de puntos x para los cuales se tiene $f(x) \geq \varepsilon$, de manera que si x_0 es un número irracional en el intervalo $[0,1]$, podemos tomar una vecindad de x_0 que no contenga a ninguno de los puntos en donde f es mayor o igual a ε. Para cualquier x en esa vecindad se tiene $f(x) - f(x_0) < \varepsilon$.

Por otra parte, f satisface el criterio R_2 de Riemann y, por lo tanto, es integrable. En efecto, dada $\sigma > 0$ y $\varepsilon > 0$, sea $A = \{x_1, x_2, \ldots, x_M\}$ el conjunto de puntos en los cuales f es mayor o igual que σ y definamos $\delta = \frac{\varepsilon}{3M}$. Si P es una partición de norma menor que δ, hay a lo más $2M$ subintervalos de P que contienen algún punto de A; en el resto de los subintervalos de P la oscilación de f es menor que σ, de manera que si $\lambda(P, \sigma)$ es la suma de las longitudes de los subintervalos de P en los cuales la oscilación de la función es mayor que σ, se tiene $\lambda(P, \sigma) \leq 2M\delta < \varepsilon$, así que $\lim_{\|P\| \to 0} \lambda(P, \sigma) = 0$.

Hermann Hankel, discípulo de Riemann, introdujo en 1870 ([**43**]) el concepto de oscilación de una función en un punto y reformuló el criterio de Riemann en los siguientes términos:

> Sea $f : [a, b] \to \mathbb{R}$ una función acotada y $x \in (a, b)$. Sea $(I_n)_{n \in \mathbb{N}}$ una sucesión de intervalos cerrados encajados que contengan a x como punto interior y tales que $\bigcap_{n=1}^{\infty} I_n = \{x\}$; denotemos por O_n a la oscilación de f en el intervalo I_n; entonces el límite $\lim_{n \rightsquigarrow \infty} O_n$ existe y es independiente de la sucesión particular de intervalos encajados con las propiedades dadas antes. A ese límite se le llama la oscilación de la función f en el punto x.

Demostró entonces, erróneamente, que una función es integrable si y sólo si para cualquier $\varepsilon > 0$ el conjunto de puntos donde la oscilación de la función es mayor que ε es denso en ninguna parte.

Durante varios años prevaleció la búsqueda de la caracterización de las funciones integrables en base a la pequeñez topológica del conjunto de sus discontinuidades y, en esa búsqueda, se puede observar la confusión que existía respecto a los diferentes conceptos de pequeñez que podían definirse.

Alrededor del año 1873 tal confusión radicaba básicamente en la idea de que un conjunto es denso en ninguna parte si y sólo si es de primera especie.

Recordemos que si $A \subset \mathbb{R}$, se denota por $A^{(1)}$ al conjunto de puntos de acumulación de A, por $A^{(2)}$ al conjunto de puntos de acumulación de $A^{(1)}$, etc... Al conjunto $A^{(n)}$ se le llama el enésimo conjunto derivado de A. Se dice que un conjunto $A \subset \mathbb{R}$ es de primera especie si $A^{(n)}$ es finito para alguna n.

En 1873 era ya bien conocido que un conjunto acotado de primera especie es denso en ninguna parte: si un conjunto es denso en algún intervalo, entonces el conjunto de sus puntos de acumulación también lo es; de manera que ese conjunto no puede ser de primera especie.

Sin embargo, se pensaba que los conjuntos de primera especie agotaban las posibilidades de los conjuntos densos en ninguna parte. La confusión terminó cuando se inventaron métodos para construir conjuntos densos en ninguna parte.

Paul du Bois Reymond dio en 1883 ([**30**]) un ejemplo de un conjunto denso en ninguna parte que no es de primera especie:

Sea I_n una sucesión de intervalos ajenos cuyos puntos extremos convergen al punto P.

En el interior de I_n definamos un conjunto Q_n de orden n y sea $Q = \bigcup_{n=1}^{\infty} Q_n$.

Q es un conjunto denso en ninguna parte pues cada conjunto Q_n lo es y éstos se encuentran en intervalos ajenos.

Por otra parte, $P \in Q^{(n)}$ para toda n, por lo tanto, Q no es de primera especie.

Otro método de construcción de conjuntos densos en ninguna parte fue desarrollado de manera independiente por Henry John Stephen Smith en 1875 ([**84**]), Vito Volterra en 1881 ([**92**], [**93**]) y Georg Ferdinand Ludwig Philipp Cantor durante el periodo 1879-1884 ([**11**], [**12**], [**13**], [**14**], [**15**]). Este método es el que se utiliza actualmente para definir el conjunto de Cantor, el cual es un ejemplo de un conjunto denso en ninguna parte que no es de primera especie.

Definamos:

$F_0 = [0, 1]$,

$F_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$,

$F_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$,

$\vdots$

En general, si ya tenemos definido el conjunto F_n, éste consta de una unión de 2^n intervalos cerrados ajenos. El conjunto F_{n+1} se construye entonces partiendo cada uno de esos intervalos en 3 intervalos de la misma longitud y eliminando el intervalo central abierto.

$F = \bigcap_{n=1}^{\infty} F_n$ es llamado el conjunto de Cantor y tiene las siguientes propiedades:

- Es un conjunto denso en ninguna parte
- $F = F^{(n)}$ para toda n, por lo tanto, no es de primera especie.

Durante ese periodo emergió una nueva clase de conjuntos, los de contenido cero:

DEFINICIÓN 1.2. *Se dice que un conjunto tiene contenido cero si, para cualquier $\varepsilon > 0$, existe una familia finita de intervalos abiertos cuya unión cubre al conjunto y tales que la suma de sus longitudes es menor que ε.*

Se pudo demostrar además que esta nueva clase de conjuntos se ubica entre las otras dos que hemos mencionado, es decir, todo conjunto acotado de primera especie tiene contenido cero y a su vez todo conjunto de contenido cero es denso en ninguna parte.

> Una demostración de la primera de estas contenciones puede basarse en el hecho de que si B es un conjunto acotado tal que el conjunto de sus puntos de acumulación tiene contenido cero entonces B también tiene contenido cero. En efecto, sea C el conjunto de puntos de acumulación de B; entonces, dada $\varepsilon > 0$, existe una colección finita de intervalos abiertos cuya unión cubre a C y tales que la suma de sus longitudes es menor que $\frac{\varepsilon}{2}$; el conjunto de puntos de B que no son cubiertos por la unión de esos intervalos es finito, pues si fuera infinito, siendo además acotado, tendría por lo menos un punto de acumulación, el cual obviamente no estaría en C, lo cual es una contradicción. Tal conjunto finito puede ser cubierto por una colección finita de intervalos abiertos tales que la suma de sus longitudes es menor que $\frac{\varepsilon}{2}$.
>
> Para demostrar la segunda contención, sea B un conjunto denso en algún intervalo $[a, b]$ con $a < b$; entonces, dada cualquier colección finita de intervalos abiertos cuya unión cubra a B, redefínanse los intervalos de tal manera que se tengan intervalos ajenos con la misma unión; de esta forma resulta fácil mostrar que el conjunto de puntos del intervalo $[a, b]$ que no son cubiertos por la unión de esos intervalos es finito, pues de otra manera existiría un intervalo no vacío (c, d), contenido en $[a, b]$, el cual no tendría puntos en común con la unión de tales intervalos; pero, como B es denso en $[a, b]$, el intervalo (c, d) contendría puntos de B, lo cual es una contradicción. Se concluye entonces que dada cualquier colección finita de intervalos abiertos cuya unión cubra a B, la suma de las longitudes de esos intervalos es mayor o igual a $b - a$, de manera que B no puede tener contenido cero.

El conjunto de Cantor, además de ser denso en ninguna parte, tiene contenido cero y, como ya se mencionó, no es de primera especie. Además, con el mismo método con el que se construye el conjunto de Cantor, se pudieron construir conjuntos compactos, densos en ninguna parte, que no son de primera especie y, además, que no tienen contenido cero. Por ejemplo, divídase el intervalo $[0, 1]$ en 3 intervalos de la misma longitud, elimínese el interior del subintervalo central y llámese F_1 a la unión de los subintervalos cerrados que restan; divídase cada uno de los 2 subintervalos que forman F_1 en 3^2 intervalos de la misma longitud, elimínese el interior del subintervalo central, llámese F_2 a la unión de los subintervalos cerrados que restan; júntese cada grupo de subintervalos contiguos en un solo intervalo, divídase cada uno de los 2^2 subintervalos que se forman en 3^3 intervalos de la misma longitud y elimínese el interior del subintervalo central; continuando con este proceso indefinidamente, la intersección F de los conjuntos $F_1, F_2, \ldots$ resulta ser un conjunto compacto, denso en ninguna parte y que no tiene contenido cero.

Para probar que F es denso en ninguna parte, obsérvese que cada uno de los 2^n intervalos ajenos que forman F_n tiene longitud igual a:

$\left(\frac{1}{3}\frac{3-1}{2}\right)\left(\frac{1}{3^2}\frac{3^2-1}{2}\right)\cdots\left(\frac{1}{3^n}\frac{3^n-1}{2}\right) = \frac{1}{2^n}\frac{3-1}{3}\frac{3^2-1}{3^3}\cdots\frac{3^n-1}{3^n} < \frac{1}{2^n}.$

Supongamos que F es denso en algún intervalo (a,b) con $0 \leq a < b \leq 1$, entonces, como F es cerrado, se tiene $[a,b] \subset F$ y, por lo tanto, $[a,b] \subset F_n$ para cualquier n. Así que, como F_n es la unión de 2^n intervalos ajenos, se tiene $[a,b] \subset I$, donde I es alguno de los 2^n intervalos ajenos que componen F_n. De manera que $b - a \leq l(I) < \frac{1}{2^n}$. Como esto pasa para cualquier $n \in \mathbb{N}$, se concluye que $b - a = 0$, lo cual es una contradicción. Por lo tanto, F es denso en ninguna parte.

Ahora bien, al dividir cada uno de los 2^n intervalos ajenos que forman F_n en 3^{n+1} subintervalos de la misma longitud y eliminar el subintervalo central, la suma de las longitudes de los intervalos que se eliminan está dada por:

$2^n\frac{1}{3^{n+1}}\left(\frac{1}{3}\frac{3-1}{2}\right)\left(\frac{1}{3^2}\frac{3^2-1}{2}\right)\cdots\left(\frac{1}{3^n}\frac{3^n-1}{2}\right) = \frac{1}{3^{n+1}}\frac{3-1}{3}\frac{3^2-1}{3^3}\cdots\frac{3^n-1}{3^n} < \frac{1}{3^{n+1}}.$

De manera que si S es la suma de las longitudes de todos los intervalos abiertos ajenos que componen F^c, se tiene:

$S < \frac{1}{3} + \frac{1}{3^2} + \cdots = \frac{1}{2}.$

Sea $I_1, I_2, \ldots, I_n$ una colección finita de intervalos abiertos cuya unión cubre F, entonces esos intervalos, junto con los intervalos abiertos ajenos que componen F^c, forman una cubierta del intervalo $[0,1]$. Por lo tanto $S + \sum_{j=1}^{n} l(I_j) \geq 1$, así que:

$\sum_{j=1}^{n} l(I_j) \geq 1 - S > \frac{1}{2}.$

Se concluye entonces que F no puede tener contenido cero.

El ejemplo de Du Bois Reymond aún no era suficiente para aclarar la diferencia entre los 3 tipos de conjuntos, pues el conjunto que él definió es de contenido cero. En efecto, cada conjunto Q_n tiene contenido cero por ser de primera especie y al cubrir el punto P con cualquier intervalo abierto quedan cubiertos todos los conjuntos Q_n a partir de una cierta n.

Todo lo anterior permitió exhibir funciones no integrables cuyo conjunto de discontinuidades sea denso en ninguna parte:

El conjunto de discontinuidades de la función indicadora de un conjunto F, denso en ninguna parte, es el conjunto F. En efecto, sea F un conjunto denso en ninguna parte y sea $f : [0,1] \to \mathbb{R}$ su función indicadora.

Si $x_0 \in F^c$, existe entonces un intervalo abierto I tal que $x_0 \in I \subset F^c$, de manera que $f(x) = 0$ para cualquier $x \in I$. Por lo tanto, f es continua en x_0.

Si $x_0 \in F$, entonces, como F es denso en ninguna parte, para cualquier intervalo abierto I que contenga a x_0, se tiene $I \cap F^c \neq \phi$, de manera que existe $y \in I$ tal que $f(y) = 0$. Por lo tanto, f no es continua en x_0.

Finalmente, la función indicadora de un conjunto F, denso en ninguna parte, que no tiene contenido cero, no es integrable.

En efecto, sea $F \subset [0,1]$ un conjunto denso en ninguna parte que no tiene contenido cero y sea $f : [0,1] \to \mathbb{R}$ la función indicadora de F. Consideremos una partición $\{x_0, x_1, \ldots, x_n\}$ del intervalo $[0,1]$.

Como F es denso en ninguna parte, cada subintervalo de la partición contiene una intervalo que no contiene puntos de F, de manera que, cualquiera que sea la partición, se puede elegir en cada subintervalo $[x_{k-1}, x_k]$ un punto ξ_k que no pertenece a F y entonces la suma $\sum_{k=1}^{n} f(\xi_k)\,(x_k - x_{k-1})$ es igual a cero.

Por otra parte, como F no tiene contenido cero, existe $\varepsilon_0 > 0$ tal que, dada cualquier colección finita de intervalos abiertos $I_1, I_2, \ldots, I_m$ cuya unión contenga a F, la suma de las longitudes de esos intervalos es mayor o igual que ε_0. Si en lugar de intervalos abiertos, consideramos una colección finita de intervalos cerrados $I_1, I_2, \ldots, I_m$ cuya unión contenga a F, entonces, dada cualquier $\delta > 0$, podemos cubrir los extremos de esos intervalos con $2m$ intervalos abiertos $J_1, J_2, \ldots, J_{2m}$ tales que la suma de sus longitudes es menor que δ. Por lo tanto, como el interior de los intervalos $I_1, I_2, \ldots, I_m$ junto con los intervalos $J_1, J_2, \ldots, J_{2m}$ cubren a F, la suma de sus longitudes es mayor o igual que ε_0. Se tiene entonces:

$\sum_{k=1}^{m} l\,(I_k) \geq \varepsilon_0 - \sum_{k=1}^{2m} l\,(J_k) > \varepsilon_0 - \delta.$

Como esto es válido para cualquier $\delta > 0$, se concluye que $\sum_{k=1}^{m} l\,(I_k) \geq \varepsilon_0$.

Dada cualquier partición $\{x_0, x_1, \ldots, x_n\}$ del intervalo $[0,1]$, consideremos los subintervalos $[x_{k-1}, x_k]$ que contengan por lo menos un elemento de F. En cada uno de esos subintervalos tomemos un punto ξ_k que pertenece a F y en el resto tomemos un punto ξ_k que no pertenece a F. Entonces la suma $\sum_{k=1}^{n} f(\xi_k)\,(x_k - x_{k-1})$ es mayor o igual que ε_0.

Así que eligiendo los puntos ξ_k de la primera manera obtenemos una suma

$\sum_{k=1}^{n} f(\xi_k)\,(x_k - x_{k-1})$

igual a cero y eligiendo los puntos ξ_k de la segunda manera obtenemos una suma

$\sum_{k=1}^{n} f(\xi_k)\,(x_k - x_{k-1})$

mayor o igualque ε_0.

Por lo tanto, f no es integrable.

Por otro lado, Riemann había mostrado la existencia de funciones cuyas discontinuidades forman un conjunto denso pero que son integrables. Se podía concluir, finalmente, que:

no es el tamaño topológico del conjunto de discontinuidades lo que determina que una función sea o no sea integrable.

Fue en ese momento cuando se pudo ya establecer con toda claridad la condición para que una función sea integrable. Axel Harnack demostró en 1881 ([**44**]) que:

Una función es integrable si y sólo si, para cualquier $\boldsymbol{\sigma} > 0$, el conjunto de puntos donde la oscilación de la función es mayor que $\boldsymbol{\sigma}$ tiene contenido cero.

La demostración es como sigue:

Sea $f : [a, b] \to \mathbb{R}$ una función acotada. Obsérvese primero que si P es una partición del intervalo $[a, b]$ y, para $\sigma > 0$ dada, x es un punto en donde la oscilación de f es mayor que σ entonces o bien x pertenece al interior de uno de los subintervalos de P en donde la oscilación de f es mayor que σ, o bien x es un elemento de la partición P. Sea entonces $A(\sigma)$ el conjunto de puntos $x \in [a, b]$ donde la oscilación de f es mayor que σ.

Si f es Riemann integrable, satisface el criterio R_2, así que, dada $\varepsilon > 0$, existe $\delta > 0$ tal que si P es una partición de norma menor que δ, entonces $\lambda(P, \sigma) < \frac{\varepsilon}{2}$; tomemos entonces una partición particular P de norma menor que δ e intervalos abiertos $I_1, I_2, \ldots, I_n$ cuya unión cubra los puntos de la partición P y tales que la suma de sus longitudes sea menor que $\frac{\varepsilon}{2}$. De esta manera, $A(\sigma)$ queda cubierto por una colección finita de intervalos abiertos tales que la suma de sus longitudes es menor que ε, así que $A(\sigma)$ tiene contenido cero.

Supongamos ahora que $A(\sigma)$ tiene contenido cero para cualquier $\sigma > 0$ y sea $\sigma > 0$ fija. Dada $\varepsilon > 0$, existen entonces intervalos abiertos no vacíos, I_k $(k = 1, 2, \ldots, n)$, de extremos $c_k, d_k \in [a, b]$, tales que $d_k \leq c_{k+1}$ para $k = 1, 2, \ldots, n-1$, $A(\sigma) \subset \cup_{k=1}^{n} I_k$ y $\sum_{k=1}^{n} [d_k - c_k] < \varepsilon$. Dado un intervalo I de la forma $[a, c_1]$, $[d_k, c_{k+1}]$ $(k = 1, 2, \ldots, n-1)$ o $[d_n, b]$, si la oscilación de f en I es mayor que 2σ, se parte I en dos subintervalos de la misma longitud; si, en ninguno de los subintervalos que se forman, la oscilación de f sigue siendo mayor que 2σ, termina el proceso; si no, se parte en dos subintervalos de la misma longitud cada subintervalo en donde la oscilación de f sea mayor que 2σ.

Continuando con este proceso, en un número finito de pasos se tiene partido I de tal manera que en cada subintervalo de la partición la oscilación de f es menor o igual a 2σ. De esta manera se tiene una partición del intervalo $[a,b]$ formada por todos los extremos de los intervalos I_k $(k = 1, 2, , \ldots, n)$ y por todos los extremos de los subintervalos que conforman cada uno de los intervalos I del tipo descrito arriba.

Definamos entonces δ_1 como la más pequeña de las longitudes de los subintervalos de esta partición y $\delta = \text{mín}\{\delta_1, \frac{\varepsilon}{n}\}$. Dada una partición P de norma menor que δ, si la oscilación de f es mayor que 4σ en un subintervalo de P, entonces ese subintervalo intersecta alguno de los intervalos I_k $(k = 1, 2, , \ldots, n)$, por lo tanto, se tiene $\lambda(P, 4\sigma) < \varepsilon + 2n\frac{\varepsilon}{n} = 3\varepsilon$. Así que, por el criterio R_2, f es integrable.

El concepto de contenido cero se convertiría desde ese momento en uno clave para la teoría de la integración. Surgiría más adelante en conexión con la teoría de integrales dobles sobre una región E del plano, cuya frontera requiere tener contenido cero para que la integral pueda ser definida.

En ese momento se tuvieron entonces las bases para desarrollar una teoría del contenido, lo cual fue llevado a cabo por Otto Stolz ([**88**], [**89**]), Axel Harnack ([**45**], [**46**]), Giuseppe Peano ([**74**]) y, sobre todo, por Marie Ennemond Camille Jordan ([**50**]). Todo esto durante el periodo que va de 1883 a 1892:

Las definiciones y propiedades se establecieron en ese periodo tanto para el caso de subconjuntos de los reales como para subconjuntos del plano, siendo similares en los dos casos. También surgieron en este periodo los conceptos de integral superior e inferior de una función, las cuales serán denotadas en lo que sigue por $\overline{\int}$ e $\underline{\int}$ respectivamente.

Sea A un conjunto acotado de números reales y $[a,b]$ un intervalo que lo contenga. Para cada partición P del intervalo $[a,b]$ sea $\bar{S}(P, A)$ la suma de los subintervalos de P que contienen puntos de A y $\underline{S}(P, A)$ la suma de los subintervalos de P contenidos en A. Se define entonces el **contenido exterior** de A, $c_e(A)$, y el **contenido interior** de A, $c_i(A)$ mediante las relaciones:

$$c_e(A) = \text{ínf}\left\{\bar{S}(P, A) : P \text{ es partición del intervalo } [a,b]\right\}.$$

$$c_i(A) = \text{ínf}\{\underline{S}(P, A) : P \text{ es partición del intervalo } [a,b]\}.$$

Se dice entonces que A es **Jordan-medible** si $c_e(A) = c_i(A)$ y, en este caso, a esta cantidad común se le llama el **contenido** de A y se le denota por $c(A)$.

Evidentemente todo conjunto de contenido cero es Jordan-medible. También todo intervalo acotado es Jordan-medible y su contenido es igual a su longitud.

Consideremos un intervalo $[a, b]$, entonces la familia de subconjuntos de $[a, b]$ que son Jordan-medibles es cerrada bajo complementos y uniones finitas. Además, si A_1, A_2, …,A_n es una familia finita de subconjuntos de $[a, b]$ que son Jordan-medibles y ajenos por parejas, entonces:

$$c\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} c\left(A_k\right).$$

Se observó también durante ese periodo que la teoría del contenido está íntimamente relacionada con la teoría de integración de Riemann, no únicamente porque la caracterización de la integrabilidad de una función se establece con base en el concepto de contenido cero o porque para integrar sobre una región del plano se requiere que la frontera de ésta tenga contenido cero. La relación resulta bastante más profunda, a tal grado que puede decirse que constituyen en realidad la misma teoría, formulada por un lado para los conjuntos y por el otro para las funciones. Por ejemplo, se tienen los siguientes resultados:

Proposición 1.1. *Sea A un subconjunto del intervalo $[a, b]$ e I_A su función indicadora, entonces:*

$$\overline{\int_a^b} I_A(x)dx = c_e(A),$$

$$\underline{\int_a^b} I_A(x)dx = c_i(A).$$

Corolario 1.1. *A es Jordan-medible si y sólo si I_A es Riemann integrable. Además, en ese caso, se tiene:*

$$\int_a^b I_A(x)dx = c(A).$$

Proposición 1.2. *Sea $f : [a, b] \to \mathbb{R}$ una función acotada no negativa y E la región en $\mathbb{R}^2$ acotada por el eje x y la gráfica de f entre a y b, entonces f es Riemann integrable si y sólo si E es Jordan medible. Además, en ese caso, se tiene $\int_a^b f(x)dx = c(E)$.*

Proposición 1.3. *Sea $f : E \to \mathbb{R}$ una función acotada definida sobre un subconjunto acotado de $\mathbb{R}$ o de $\mathbb{R}^2$ Jordan medible. Para cada partición P de E en n conjuntos $E_1, E_2, \ldots, E_n$ Jordan medibles ajenos, definamos $\bar{S}(P, f) = \sum_{j=1}^{n} M_j c(E_j)$ y $\underline{S}(P, f) = \sum_{j=1}^{n} m_j c(E_j)$, donde, para cada $j \in \{1, 2, \ldots, n\}$, $M_j = \sup\{f(x) : x \in E_j\}$ y $m_j = \text{ínf}\{f(x) : x \in E_j\}$, entonces:*

$$\overline{\int}_E f(x)dx = \text{lím}_{\|P\|\to 0} \bar{S}(P, f),$$

$$\underline{\int}_E f(x)dx = \text{lím}_{\|P\|\to 0} \underline{S}(P, f),$$

donde $\|P\| = \text{máx}\{C(E_j) : j \in \{1, 2, \ldots, n\}\}$.

En particular, f es Riemann integrable sobre E si y sólo si el límite de las sumatorias $\sum_{j=1}^{n} f(\xi_j)c(E_j)$, donde, para cada $j \in \{1, 2, \ldots, n\}$, $\xi_j \in E_j$, existe cuando $\|P\|$ tiende a 0 y es independiente de los puntos $\xi_j \in E_j$ que se tomen. Además, en ese caso, se tiene:

$$\int_E f(x)dx = \lim_{\|P\|\to 0} \sum_{j=1}^{n} f(\xi_j)c(E_j).$$

1.2. Teoría de la medida de Borel

En 1894-1895, Félix Édouard Justin Émile Borel ([**5**], [**6**]) dio las bases para un nuevo avance al introducir el concepto de medida cero:

DEFINICIÓN 1.3. *Se dice que un conjunto tiene medida cero si para cualquier $\varepsilon > 0$ existe una colección numerable de intervalos abiertos $\{I_n\}$ cuya unión cubre al conjunto y tales que la suma de sus longitudes es menor que ε.*

Curiosamente, el concepto de medida cero no lo introdujo Borel con relación a la teoría de integración. Al introducir ese concepto, Borel estaba atacando un problema de continuación analítica de una función de variable compleja:

Considérese la función de variable compleja

$$f(z) = \sum_{n=1}^{\infty} \frac{A_n}{z-a_n},$$

donde $A_1, A_2, \ldots$ son números complejos tales que la serie $\sum_{n=1}^{\infty} |A_n|$ converge y $a_1, a_2, \ldots$ son puntos en el plano complejo que están sobre una curva cerrada C formando un conjunto denso en esa curva.

Se puede ver inmediatamente que si $z \notin C$ entonces la serie $\sum_{n=1}^{\infty} \frac{A_n}{z-a_n}$ converge pues la distancia de z a C es positiva. Consideremos dos puntos P y Q, el primero al interior de la región que forma C y el segundo al exterior de la misma; el problema que se planteó Borel consiste entonces en encontrar un arco circular que una P con Q sobre el cual la serie $\sum_{n=1}^{\infty} \frac{A_n}{z-a_n}$ converja absoluta y uniformemente. Esto llevó a Borel a la necesidad de demostrar que existen puntos z sobre C para los cuales la serie en consideración converge.

Para simplificar el razonamiento, consideremos el mismo problema pero con funciones de variable real.

Sea $\{a_1, a_2, \ldots\}$ un conjunto numerable y denso en el intervalo $[a, b]$ y $(A_n)_{n\geq 1}$ una sucesión de números reales. Para cada $x \in [a, b] - \{a_1, a_2, \ldots\}$ consideremos la serie $\sum_{n=1}^{\infty} \frac{A_n}{x-a_n}$. Aparentemente tal serie no converge para ninguno de esos puntos x pues el conjunto $\{a_1, a_2, \ldots\}$ es denso en $[a, b]$ y entonces dado cualquier punto $x \in [a, b]$ hay puntos a_n tan cerca de x como se quiera. Sin embargo, siguiendo a Borel, se puede mostrar que, asumiendo que la serie $\sum_{n=1}^{\infty} \sqrt{|A_n|}$ converge, existe una infinidad no numerable de puntos $x \in [a, b]$ para los cuales la serie converge. En efecto, para cada $n \in \mathbb{N}$, sea $u_n = \sqrt{|A_n|}$. Sea ahora l la longitud del intervalo $[a, b]$ y $N \in \mathbb{N}$ tal que $\sum_{n=N+1}^{\infty} u_n < \frac{l}{2}$. Para cada $n > N$ sea I_n un intervalo abierto con centro en a_n y radio u_n. Se tiene entonces $\sum_{n=N+1}^{\infty} l(I_n) < l$, donde $l(I_n)$ es la longitud del intervalo I_n. Como los puntos $a_1, a_2, \ldots, a_N$ forman un conjunto finito, se pueden cubrir con intervalos abiertos $I_1, I_2, \ldots, I_n$, respectivamente, de tal manera que $\sum_{n=1}^{\infty} l(I_n) < l$. Si x no pertenece a ninguno de los intervalos $I_{N+1}, I_{N+2}, \ldots$ entonces $|x - a_i| > 0$ para $i \in \{1, 2, \ldots, N\}$ y $|x - a_i| \geq u_i$ para $i \in \{N + 1, N + 2, \ldots\}$. Por lo tanto:
$\sum_{n=1}^{\infty} \left|\frac{A_n}{x-a_n}\right| = \sum_{n=1}^{N} \left|\frac{A_n}{x-a_n}\right| + \sum_{n=N+1}^{\infty} \left|\frac{A_n}{x-a_n}\right|$
$\leq \sum_{n=1}^{N} \left|\frac{A_n}{x-a_n}\right| + \sum_{n=N+1}^{\infty} \sqrt{|A_n|} < \infty.$

Lo único que resta probar es que existe una infinidad de puntos $x \in [a, b]$ que no pertenecen a ninguno de los intervalos $I_{N+1}, I_{N+2}, \ldots$. Para esto, Borel demostró el resultado, ahora clásico, que asegura que todo intervalo cerrado y acotado es compacto. De manera más específica, Borel demostró, básicamente como se hace actualmente, que si un intervalo cerrado y acotado es cubierto por una infinidad numerable de intervalos abiertos, entonces existe una colección finita de esos intervalos que también lo cubren.

Con base en ese resultado, si los intervalos $I_1, I_2, \ldots$ cubrieran al intervalo $[a, b]$, necesariamente se tendría $\sum_{n=1}^{\infty} l(I_n) \geq l$, lo cual es una contradicción. Más aún, si únicamente hubiera una colección numerable de puntos $x \in [a, b]$ que no pertenecen a ninguno de los intervalos $I_1, I_2, \ldots$., estos puntos podrían ser cubiertos por una nueva colección numerable de intervalos abiertos de tal manera que la suma de sus longitudes, sumadas con las longitudes de los intervalos $I_1, I_2, \ldots$, siga siendo menor que l, lo cual no es posible.

> Todavía siguiendo a Borel, se puede decir aún más, pues cambiando l por una $\varepsilon > 0$ arbitraria en el razonamiento anterior se muestra que el conjunto de puntos $x \in [a, b]$ para los cuales la serie $\sum_{n=1}^{\infty} \frac{A_n}{x-a_n}$ no converge absolutamente pueden ser cubiertos por una colección numerable de intervalos abiertos de tal manera que la suma de sus longitudes sea menor que ε. Es decir, utilizando el concepto que introdujo Borel, el conjunto de puntos $x \in [a, b]$ para los cuales la serie $\sum_{n=1}^{\infty} \frac{A_n}{x-a_n}$ no converge absolutamente tiene medida cero.

Además de introducir el concepto de medida cero al resolver el problema que se planteó, la demostración de Borel contiene un resultado que sería clave para que más adelante Lebesgue. pudiera definir el concepto de medida. Ese resultado se puede enunciar de la siguiente manera:

Sea I un intervalo cerrado y acotado y $(I_j)_{j\in\mathbb{N}}$ una sucesión de intervalos abiertos tales que $I \subset \bigcup_{j=1}^{\infty} I_j$, entonces:

$$l(I) \leq \sum_{j=1}^{\infty} l(I_j).$$

El resultado parece trivial ya que al evaluar la suma de las longitudes de los intervalos I_j, si dos de ellos se traslapan, podría haber una parte de I cuya longitud se está sumando dos veces; si no se traslapan, al sumar las longitudes de los dos intervalos, esa suma es por lo menos igual a la suma de las longitudes de las partes de I que se encuentran dentro de esos intervalos. Sin embargo, al tratar de formalizar esta idea se llega nuevamente al problema inicial.

Si el conjunto de intervalos abiertos cuya unión cubre I fuera finito, el resultado se puede demostrar fácilmente. En efecto, supongamos que el intervalo acotado $I = [a, b]$ está contenido en la unión de los intervalos no vacíos $I_j = (a_j, b_j)$, con $j \in \{1, 2, \ldots, m\}$. Si alguno de esos intervalos tiene longitud infinita, el resultado es trivial, así que podemos suponer que todos los intervalos I_j son finitos. Sea $D = \{x_0, x_1, \ldots, x_n\}$ el conjunto que se obtiene al ordenar, del menor al mayor, los puntos $a, b, a_1, b_1, a_2, b_2, \ldots, a_m, b_m$.

Para $k \in \{1, 2, \ldots, n\}$ definamos $I^{(k)} = (x_{k-1}, x_k)$ e $\overline{I}^{(k)} = [x_{k-1}, x_k]$; además, para $j \in \{1, 2, \ldots, m\}$, denotemos por $\overline{I}_j$ al intervalo $[a_j, b_j]$. Entonces: 1) los intervalos $I^{(1)}, I^{(2)}, \ldots, I^{(n)}$ son ajenos por parejas; 2) el intervalo I, así como cada uno de los intervalos $\overline{I}^{(j)}$, con $j \in \{1, 2, \ldots, m\}$, es la unión de algunos de los intervalos $\overline{I}_k$, con $k \{1, 2, \ldots, n\}$.

Sean D el conjunto de superíndices de los intervalos de la familia $\left\{\overline{I}^{(k)} : k \in \{1, 2, \ldots, n\}\right\}$ cuya unión es igual a I. Como $I \subset \bigcup_{j=1}^{m} I_j$, cada intervalo $\overline{I}^{(k)}$, con $k \in D$, está contenido en algún rectángulo $\overline{I}_j$ $(j \in \{1, 2, \ldots, m\})$. Además:

$l(I) = \sum_{\{k\in D\}} l\left(\overline{I}^{(k)}\right)$.

Para cada $j \in \{1, 2, \ldots, m\}$, sea D_j el conjunto de superíndices de los intervalos de la familia $\left\{\overline{I}^{(k)} : k \in D\right\}$ que están contenidos en $\overline{I}_j$. Obviamente se tiene:

$l(I_j) = l\left(\overline{I}_j\right) \geq \sum_{\{k \in D_j\}} l\left(\overline{I}^{(k)}\right)$.

Y como cada intervalo $\overline{I}^{(k)}$, con $k \in D$, está contenido en algún rectángulo $\overline{I}_j$ ($j \in \{1, 2, \ldots, m\}$), se tiene:

$\sum_{j=1}^{m} l(I_j) \geq \sum_{j=1}^{m} \sum_{\{k \in D_j\}} l\left(\overline{I}^{(k)}\right) \geq \sum_{\{k \in D\}} l\left(\overline{I}^{(k)}\right) = l(I)$.

Regresando al problema inicial, sea I un intervalo cerrado y acotado e $(I_j)_{j \in \mathbb{N}}$ una sucesión de intervalos abiertos tales que $I \subset \bigcup_{j=1}^{\infty} I_j$. Para demostrar el resultado enunciado, Borel demostró que existe una colección finita de los intervalos I_j cuya unión también contiene a I. Su razonamiento, ahora clásico, fue el siguiente:

Sea

$B = \{x \in [a, b]$: el intervalo $[a, x]$ está contenido en la unión de un número finito

de intervalos de la familia $\{I_j : j \in \mathbb{N}\}\}$.

B es un conjunto no vacío, ya que $a \in B$, y está acotado por b; por lo tanto B tiene un supremo, que denotaremos por x_0.

Como $a \in B$ y b es cota superior de B se tiene $a \leq x_0 \leq b$.

Como $x_0 \in [a, b]$, hay un intervalo $I_{j_0} = (a_{j_0}, b_{j_0})$ al cual pertenece x_0. Siendo x_0 el supremo de B, existe $x \in B$ tal que $x \in (a_{j_0}, x_0]$. Como $x \in B$, hay una colección finita de intervalos de la familia $\{I_j : j \in \mathbb{N}\}$ cuya unión cubre al intervalo $[a, x]$. Entonces agregándole a esa colección el intervalo I_{j_0} (si no está ya incluido), obtenemos una colección finita de intervalos de la familia $\{I_j : j \in \mathbb{N}\}$ cuya unión cubre al intervalo $[a, x_0]$; así que $x_0 \in B$.

Sea $I_{j_0}, I_{j_1}, I_{j_2}, \ldots, I_{j_m}$ una colección finita de intervalos cuya unión cubre el intervalo $[a, x_0]$. Si se tuviera $x_0 < b$, entonces, si $c = \text{mín}(b_{j_0}, b)$, se tendría $x_0 < c \leq b$. Tomando cualquier $y \in (x_0, c)$ se tendría $x_0 < y < b_{j_0}$, así que la unión de los intervalos $I_{j_0}, I_{j_1}, I_{j_2}, \ldots, I_{j_m}$ también cubriría al intervalo $[a, y]$. Como $(x_0, c) \subset [a, b]$, se tendría $y \in [a, b]$ y, por lo tanto, $y \in B$, lo cual no es posible ya que $x_0 < y$ y x_0 es el supremo de B. Por lo tanto, $x_0 = b$ y, entonces, el intervalo $[a, b]$ está contenido en la unión de un número finito de intervalos de la familia $\{I_j : j \in \mathbb{N}\}$.

Más adelante, en un libro publicado en 1898 ([**7**]), Borel retomó el concepto de conjunto de medida cero. para desarrollar una teoría de la medida. Para esto, influenciado en parte por el trabajo de Jules Joseph Drach, siguió el método axiomático. Para Borel la idea fundamental consistía en definir los elementos nuevos que se introducen con ayuda de sus propiedades esenciales, es decir, aquellas que son estrictamente indispensables para los razonamientos que siguen. En el caso de la medida, las propiedades esenciales que planteó Borel son las siguientes:

(i) La medida de la unión de una colección numerable de conjuntos ajenos es igual a la suma de sus medidas.
(ii) La medida de la diferencia de dos conjuntos de medida finita A y B, con $A \subset B$, es igual a la diferencia de sus medidas $m(B) - m(A)$.
(iii) La medida de un conjunto nunca es negativa.

Llamaba entonces conjuntos medibles a todos aquellos conjuntos a los cuales se les pueda asignar una medida en base a las propiedades mencionadas, tomando como punto de partida que la medida de un intervalo es su longitud.

Borel no vio relación entre su concepto de medida y el de integral. Más aún, aclaraba que el problema que él estaba investigando era totalmente diferente del resuelto por Jordan. Además, consideraba la definición que hacía Jordan de los conjuntos medibles (con contenido) como más general que la que él daba pues, por ejemplo, con base en la definición de Jordan, cualquier subconjunto del conjunto de Cantor es medible, de manera que, teniendo el conjunto de Cantor la misma cardinalidad que los números reales, la familia de conjuntos Jordan medibles tiene una cardinalidad mayor que la de los reales. Por otra parte, se puede mostrar que la familia de conjuntos medibles que define Borel tiene únicamente la cardinalidad de los números reales.

1.3. Teoría de la medida de Lebesgue

El paso siguiente en el desarrollo de la Teoría de la Medida, así como el último paso hacia la caracterización de las funciones Riemann-integrables lo dió Henri Léon Lebesgue en 1902 ([**57**])

Para la caracterización de las funciones Riemann integrables, Lebesgue primero demostró una forma ligeramente distinta del resultado de Harnack:

Si, dada $\boldsymbol{\sigma} > 0$, B($\boldsymbol{\sigma}$) denota al conjunto de puntos en donde la oscilación de la función f es mayor o igual que $\boldsymbol{\sigma}$, entonces f es integrable si y sólo si para cualquier $\boldsymbol{\sigma} > 0$, B($\boldsymbol{\sigma}$) tiene contenido cero.

Mostró además que, para cualquier $\sigma > 0$, $B(\sigma)$ es un conjunto cerrado, de manera que, siendo acotado, es compacto.

Las demostraciones de estos resultados son como sigue:

Sea $A(\sigma)$ el conjunto de puntos en donde la oscilación de f es mayor que σ, entonces:

$A(\sigma) \subset B(\sigma) \subset A(\frac{\sigma}{2})$

Así que si f es integrable entonces $A(\frac{\sigma}{2})$ tiene contenido cero y, por lo tanto, $B(\sigma)$ también. Inversamente, si $B(\sigma)$ tiene contenido cero para cualquier $\sigma > 0$, entonces $A(\sigma)$ también.

Para probar que $B(\sigma)$ es un conjunto cerrado, sea x un punto de acumulación de $B(\sigma)$, entonces toda vecindad de x contiene puntos de $B(\sigma)$, es decir, puntos en donde la oscilación de f es mayor o igual que σ. Por lo tanto, la oscilación de f en x es también mayor o igual que σ, así que x pertenece a $B(\sigma)$.

Lebesgue observó entonces que si D es el conjunto de puntos en donde la función es discontinua, se tiene $D = \bigcup_{n=1}^{\infty} B(\frac{1}{n})$. Entonces, si f es Riemann integrable, $B(\frac{1}{n})$ tiene contenido cero para cualquier $n \in \mathbb{N}$, así que D tiene medida cero. Por otra parte, si D tiene medida cero, entonces $B(\frac{1}{n})$ tiene medida cero para cualquier $n \in \mathbb{N}$, de manera que, siendo estos conjuntos compactos, también tienen contenido cero; finalmente, dada $\sigma > 0$ arbitraria y $n > \frac{1}{\sigma}$ se tiene $B(\sigma) \subset B(\frac{1}{n})$, así que $B(\sigma)$ tiene contenido cero. Se tiene así la siguiente caracterización de las funciones Riemann integrables:

Una función acotada $\mathbf{f} : [\mathbf{a}, \mathbf{b}] \mapsto \mathbb{R}$ es Riemann integrable si y sólo si el conjunto de puntos donde la función es discontinua tiene medida cero.

Lebesgue desarrolló su teoría de la integral en su tesis doctoral titulada *Integrale, longueur, aire* ([**57**]). Más tarde la expuso en su libro *Leçons sur l'intégration et la recherche des fonctions primitives* ([**58**]).

Para definir la integral, primero desarrolló su teoría de la medida de conjuntos, pero el interés de Lebesgue estaba centrado en la definición de la integral ya que había analizado antes que las propiedades de la integral con la que se trabaje juegan un papel muy importante en algunos resultados acerca de la teoría de funciones; en particular en su libro dedica un capítulo al vínculo entre la integral y la búsqueda de la primitiva de una función, es decir, una función tal que su derivada sea la función dada. Se planteó entonces el encontrar una definición de la integral con mejores propiedades que las integrales conocidas, en particular la integral de Riemann. Se propuso así asignar a cualquier función acotada f, definida en un intervalo finito (a, b), un número real, denotado por $\int_a^b f(x)\, dx$, al cual llamaba la integral de f en el intervalo (a, b). Planteó que esta integral debe tener las siguientes propiedades:

(i) Para cualesquiera $a, b, h \in \mathbb{R}$, se tiene:
$\int_a^b f(x)\, dx = \int_{a+h}^{b+h} f(x-h)\, dx$.
(ii) Para cualesquiera $a, b, c \in \mathbb{R}$, se tiene:
(iii) $\int_a^b f(x)\, dx + \int_a^c f(x)\, dx + \int_c^a f(x)\, dx = 0$.
(iv) $\int_a^b [f(x) + \varphi(x)]\, dx = \int_a^b f(x)\, dx + \int_a^b \varphi(x)\, dx$.
(v) Si f es no negativa y $a < b$, entonces $\int_a^b f(x)\, dx$ es no negativa.
(vi) $\int_0^1 1dx = 1$.

(vii) Si una sucesión de funciones $(f_n)_{n\in\mathbb{N}}$ es no decreciente y converge a la función f, entonces la sucesión de integrales $\left(\int_a^b f_n(x)\,dx\right)_{n\in\mathbb{N}}$ converge a la integral $\int_a^b f(x)\,dx$.

En seguida hizo algunas observaciones importantes acerca de estas propiedades que planteó para la integral:

> *"La significación, la necesidad y las consecuencias de las cinco primeras condiciones de este problema de integración son más o menos evidentes; no nos extenderemos acerca de ellas. La condición* 6 *tiene un lugar aparte. No tiene ni el mismo carácter de simplicidad que las cinco primeras, ni el mismo carácter de necesidad. Además, mientras que es fácil construir números que satisfagan a cualesquiera cuatro de las cinco primeras condiciones, sin satisfacer a las cinco, lo cual muestra que esas cinco condiciones son independientes, no se sabe si las seis condiciones del problema de integración son independientes o no."En una nota a pie de página, agrega: "La respuesta a esta pregunta importa poco para las aplicaciones, pero presenta interés desde el punto de vista de los principios. Si se demostrara que esta sexta condición es independiente de las otras cinco, cabría buscar reemplazarla por una sexta más simple y sobre todo buscar si, entre los sistemas de números que satisfacen solamente a las cinco primeras condiciones, no hay algunos tan útiles como el que va a ser estudiado."*

Años más tarde, Stefan Banach demostraría que la sexta condición que plantea Lebesgue es independiente de las primeras cinco ([**3**]).

Aclaraba Lebesgue que la definición de la integral que daba es descriptiva, es decir que la ha definido mediante las propiedades características que tiene. Se propuso entonces dar una definición constructiva equivalente a la descriptiva; es decir, enunciar las operaciones que se requieren realizar para definir la integral de una función acotada de tal manera que se satisfagan las seis condiciones de la definición descriptiva.

En seguida mostró Lebesgue que para dar una definición constructiva de la integral de cualquier función acotada, basta con hacerlo para las funciones que únicamente toman como valores 0 y 1 y, para una función f de este tipo, el problema de integración se traduce en asignar un número al conjunto de números reales $x \in (a,b)$ tales que $f(x) = 1$; de manera que entonces se planteó Lebesgue el **problema de la medida**, de conjuntos el cual consiste en asignar a cada conjunto acotado de números reales E, un número no negativo, $m(E)$, al cual llamará la medida de E, debiéndose satisfacer las siguientes propiedades:

(i) Si E es un conjunto acotado y $a \in \mathbb{R}$, entonces $m(E+a) = m(E)$.
(ii) Si $E_1, E_2, \ldots$ es una familia finita o infinita numerable de una sucesión de conjuntos, ajenos por parejas y contenidos en un conjunto acotado, entonces $m(\bigcup\limits_n E_n) = \sum_n m(E_n)$.

(iii) $m([0,1]) = 1$.

Aunque Lebesgue planteó el problema de asignar una medida a cualquier conjunto acotado de números reales, en realidad, como se ve más adelante en su razonamiento, lo que hizo fue encontrar una familia de conjuntos acotados de números reales a los cuales se les pueda asignar una medida de tal forma que se satisfagan las 3 propiedades mencionadas. Para esto, suponiendo que E es un elemento de esa familia, analizó las condiciones que debe satisfacer la medida de E para que se satisfagan las 3 propiedades; más aún, lo que hace es encontrar los conjuntos E para los cuales su medida queda únicamente determinada por esas 3 propiedades. Pero, para lograr esto, lo que hizo Lebesgue fue iniciar su razonamiento asumiendo que el problema de la medida tiene solución, es decir que se puede asignar una medida no negativa a cada subconjunto acotado de números reales. Después restringirá la familia de conjuntos medibles a aquellos cuya medida se pueda determinar de manera única. Luego de hacer esto viene el proceso inverso: partir de una definición de conjunto medible, al cual le asocia una medida, y mostrar que la familia de conjuntos así definida satisface las 3 propiedades que enunció.

Las condiciones sobre la medida implican que si $x \in \mathbb{R}$, entonces $m(\{x\}) = 0$.

En efecto, en primer lugar, si A y B son conjuntos acotados y $A \subset B$, entonces, como $B = A \cup (B - A)$, se tiene $m(B) = m(A) + m(B - A)$, así que $m(B - A) = m(B) - m(A)$ y $m(A) \leq m(B)$. Además, si m es cualquier número entero, la medida del intervalo $[m, m+1]$ debe ser igual a 1. Ahora, si x es cualquier número real, tomemos el único número entero m tal que $x \in [m, m+1)$; dada $\varepsilon > 0$, tomemos $n \in \mathbb{N}$ tal que $\frac{1}{n} < \varepsilon$ y consideremos los intervalos $\left[m, m + \frac{1}{n}\right)$, $\left[m, m + \frac{2}{n}\right)$, $\ldots$, $\left[m, m + \frac{n}{n}\right)$. Cada uno de estos intervamos tiene la misma medida y la suma de sus medidas es menor o igual a 1, así que la medida de cada uno de ellos es menor o igual a $\frac{1}{n}$. Además, como x pertenece a alguno de esos intervalos, se tiene $m(\{x\}) \leq \frac{1}{n} < \varepsilon$. Siendo ε arbitrario, se concluye que $m(\{x\}) = 0$.

También, la medida de un intervalo acotado $[a, b]$ debe de ser igual a su longitud.

En efecto, el razonamiento anterior nos lleva a que, si $n \in \mathbb{N}$, la medida de un intervalo de longitud $\frac{1}{n}$ es igual a $\frac{1}{n}$. Así que si $n, m \in \mathbb{N}$, la medida de un intervalo de longitud $\frac{m}{n}$ es igual a $\frac{m}{n}$. De aquí se sigue que la medida de un intervalo con extremos racionales es igual a su longitud. Ahora, el caso $a = b$ ya está tratado, así que tomemos $a < b$ y dado $\varepsilon > 0$, tomemos $c = \min\left\{\varepsilon, \frac{1}{4}(b - a)\right\}$ y cuatro números racionales, r, s, u y v tales que:

$a - c < r < a < u < a + c < b - c < v < b < s < b + c$.

Entonces:

$b - a - 2\varepsilon \leq b - a - 2c < v - u = m([u, v]) \leq m([a, b]) \leq m([r, s]) = s - r < b - a + 2c \leq b - a + 2\varepsilon$.

Siendo ε arbitrario, se concluye que $m([a, b]) = b - a$.

Para definir la medida de cualquier conjunto acotado, Lebesgue hizo el siguiente razonamiento:

Si E es un conjunto acotado e $I_1, I_2, \ldots$ es una colección finita o infinita numerable de intervalos, ajenos por parejas, tales que $E \subset \cup_n I_n$, entonces se debe de tener $m(E) \leq \sum_n l(I_n)$; definió entonces la **medida exterior** de E, $m_e(E)$, como el ínfimo de esas sumas, es decir:

$$m_e(E) = \text{ínf}\,\{\textstyle\sum_n l(I_n) : I_1, I_2, \ldots \text{ son intervalos ajenos por parejas y } E \subset \bigcup_n I_n\}.$$

Aquí hay un detalle que no aclaraba Lebesgue: como el problema que plantea es asignar una medida a cada conjunto acotado, se tendría que asumir que el conjunto $\cup_n I_n$ es acotado. Esto puede hacerse sin causar algún problema; en efecto si $[a, b]$ es un intervalo tal que $E \subset [a, b]$, entonces:

$\text{ínf}\,\{\sum_n l(I_n) : I_1, I_2, \ldots \text{ son intervalos ajenos por parejas y } E \subset U_n I_n\}$

$= \text{ínf}\,\{\sum_n l(I_n \cap [a, b]) : I_1, I_2, \ldots \text{ son intervalos ajenos por parejas y } E \subset U_n I_n\}.$

Ahora bien, como se tiene que $m\,([a, b]) = m\,(E) + m\,([a, b] - E)$, entonces:

$m\,(E) = m\,([a, b]) - m\,([a, b] - E) \geq m\,([a, b]) - m_e\,([a, b] - E) = l\,([a, b]) - m_e\,([a, b] - E).$

Se sigue que la cantidad:

$$l\,([a, b]) - m_e\,([a, b] - E)$$

es una cota inferior para la medida de E, la cual define como la **medida interior** de E y la denota por $m_i(E)$.

Lebesgue no demostró que la cantidad $l\,([a, b]) - m_e\,([a, b] - E)$ es la misma cualquiera que sea el intervalo $[a, b]$, conteniendo E, que se tome; sin embargo esto es cierto, así que la medida interior queda bien definida. En efecto:

Si $[a_1, b_1]$ y $[a_2, b_2]$ son intervalos que contienen a E, su intersección $[a, b]$ también lo contiene, así que para mostrar que tomando $[a_1, b_1]$ se obtiene el mismo resultado que tomando $[a_2, b_2]$, basta con demostrar que tomando $[a, b]$ se obtiene el mismo resultado que tomando cualquiera de los dos. Tomemos entonces dos intervalos, $[a, b]$ y $[c, d]$, tales que $E \subset [a, b] \subset [c, d]$ y, para cada colección finita o infinita numerable de intervalos $I_1, I_2, \ldots$, ajenos por parejas, tales que $[c, d] - E \subset \bigcup_n I_n$, definamos:

$J_n = I_n \cap [a, b]$,

$K_n = I_n \cap [c, a)$,

$L_n = I_n \cap (b, d]$.

Se tiene entonces lo siguiente:

Los intervalos $J_1, K_1, L_1, J_2, K_2, L_2, \ldots$ son ajenos por parejas y la unión de todos ellos es igual a $(\bigcup_n I_n) \cap [c,d]$.

$[a,b] - E \subset \bigcup_n J_n$.

$[c,d] - [a,b] = \bigcup_n (K_n \cup J_n)$.

Así que:

$m_e([c,d] - E) = \text{ínf}\,\{\sum_n l(I_n) : I_1, I_2, \ldots \text{ son intervalos ajenos y } [c,d] - E \subset \bigcup_n I_n\}$

$= \text{ínf}\,\{\sum_n l(I_n \cap [c,d]) : I_1, I_2, \ldots \text{ son intervalos ajenos y } [c,d] - E \subset \bigcup_n I_n\}$

$= \{\text{ínf} \sum_n l(I_n \cap [a,b]) + \sum_n l(I_n \cap [c,a)) + l(I_n \cap (b,d]) :$

$I_1, I_2, \ldots \text{ son intervalos ajenos y } [c,d] - E \subset \bigcup_n I_n\}$

$= \text{ínf}\,\{\sum_n l(I_n \cap [a,b]) + l([c,d]) - l([a,b]) :$

$I_1, I_2, \ldots \text{ son intervalos ajenos y } [c,d] - E \subset \bigcup_n I_n\}$

$= l([c,d]) - l([a,b]) + \text{ínf}\,\{\sum_n l(J_n) : J_1, J_2, \ldots \text{ son intervalos ajenos y } [a,b] - E \subset \bigcup_n J_n\}$

$= l([c,d]) - l([a,b]) + m_e([a,b] - E)$.

Así que:

$l([a,b]) - m_e([a,b] - E) = l([c,d]) - m_e([c,d] - E)$.

Como ya lo mencionamos, Lebesgue hizo lo anterior asumiendo que es posible asignarle una medida a todo conjunto acotado E, sin embargo las definiciones de medida exterior e interior son independientes de esta consideración y pueden darse para cualquier conjunto. Mostró entonces que se tienen las siguientes relaciones para cualquier conjunto acotado E:

$$c_i(E) \leq m_i(E) \leq m_e(E) \leq c_e(E).$$

Además, como se mostró arriba, de ser posible asignar una medida $m(E)$ al conjunto E, se debe de tener $m_i(E) \leq m(E) \leq m_e(E)$. Por lo tanto, la medida asignada a E será única cuando sus medidas interior y exterior coincidan. De aquí que Lebesgue estableció la siguiente definición:

Definición 1.4. *Se dice que un conjunto acotado E es medible si $m_i(E) = m_e(E)$.*

Aclaraba Lebesgue que es únicamente para estos conjuntos que se estudiará el problema de la medida, aclarando no saber siquiera si existen conjuntos que no sean medibles. Pero si existen tales conjuntos, decía que el desarrollo posterior que él hace no es suficiente para afirmar ni que el problema de la medida es posible ni que es imposible para tales conjuntos.

Este comentario de Lebesgue es importante pues lo que él hizo fue encontrar cotas más finas que las que daba Jordan para la medida de un conjunto, lo cual automáticamente amplía la familia de conjuntos a los cuales se les puede asignar una medida de manera única. En efecto, la condición $c_i(E) = c_e(E)$ permite asignar a E una única medida y esa condición implica $m_i(E) = m_e(E)$. Pero se puede cumplir la condición $m_i(E) = m_e(E)$, lo cual permite asignar una única medida a E, sin que se tenga $c_i(E) = c_e(E)$. Sin embargo, no se puede asegurar que no sea posible asignarle una medida a conjuntos para los cuales $m_i(E) < m_e(E)$. En caso de que esto fuera posible, tal vez no sería de manera única (de hecho se sabe actualmente que es posible ampliar la familia de conjuntos medibles conservando las propiedades *i* y *iii* que pide Lebesgue a la medida, pero tal extensión no es única), o tal vez se puedan encontrar cotas aún más finas que las que da Lebesgue para la medida de un conjunto y se pueda definir una medida con propiedades adicionales a las que propone Lebesgue.

Mostró Lebesgue que se tiene la siguiente propiedad:

Si $E_1, E_2, \ldots$ es una colección finita o infinita numerable de conjuntos medibles, entonces la unión de ellos, así como su intersección, es medible.

Además demostró que la familia de los conjuntos medibles satisface las 3 condiciones que planteó para la medida de los conjuntos y demostró también que si $(E_n)_{n\in\mathbb{N}}$ es una sucesión decreciente de conjuntos medibles, entonces $m\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} m(E_n)$.

Finalmente observó Lebesgue que, debido a la relación:

$$c_i(E) \leq m_i(E) \leq m_e(E) \leq c_e(E).$$

cualquier conjunto Jordan medible es también Lebesgue medible y, dado que los intervalos son medibles y la familia de conjuntos medibles tiene las propiedades enunciadas arriba, todo conjunto medible de acuerdo a la definición de Borel es también Lebesgue medible. De esta forma la teoría de la medida de Lebesgue resulta más general tanto que la de Jordan como de la de Borel y las engloba a ambas.

Años más tarde, en 1914 ([**16**]), Constantin Carathéodory expresó la condición de medibilidad de un conjunto sin introducir el concepto de medida interior. De acuerdo con la definición de Carathéodory y restringiéndonos a los conjuntos acotados, como hace Lebesgue, un conjunto acotado de números reales E es medible si y sólo si se tiene:

$$m_e(A) = m_e(A \cap E) + m_e(A \cap E^c).$$

para cualquier conjunto acotado de números reales A.

Obsérvese que, de acuerdo con la definición de Lebesgue, para que un conjunto acotado sea medible se requiere que si $[a, b]$ es un intervalo que contiene a E, entonces:

$l([a, b]) - m_e([a, b] - E) = m_i(E) = m_e(E).$

Así que la condición de medibilidad de E puede darse de la siguiente forma:

$l([a,b]) = m_e(E) + m_e([a,b] - E)$.

De manera que, si se cumple la condición de medibilidad de Carathéodory, entonces se cumple la condición de medibilidad de Lebesgue.

Por otra parte, se puede demostrar que la medida exterior satisface las siguientes propiedades:

Si A y B son conjuntos acotados tales que $A \subset B$, entonces $m_e(A) \leq m_e(B)$.

Si $A_1, A_2, \ldots$ es una colección finita o infinita numerable de conjuntos acotados cuya unión es un conjunto acotado, entonces:

$$m_e\left(\bigcup_n A_n\right) \leq \sum_n m_e\left(A_n\right).$$

Así que, en particular, la desigualdad $m_e(A) \leq m_e(A \cap E) + m_e(A \cap E^c)$ se cumple para cualquier par de conjuntos acotados E y A.

Sea ahora un conjunto acotado E, medible de acuerdo con la definición de Lebesgue, A un conjunto acotado cualquiera, $[a,b]$ un intervalo tal que $A \cup E \subset [a,b]$ e $I_1, I_2, \ldots$ una colección finita o infinita numerable de intervalos, ajenos por parejas, tales que $A \subset U_n I_n \subset [a,b]$.

De la condición $l([a,b]) = m_e(E) + m_e([a,b] - E)$, se sigue que $[a,b] - E$ es Lebesgue medible. Así que, tanto los conjuntos $I_1 \cap E, I_2 \cap E, \ldots$ como los conjuntos $I_1 \cap E^c, I_2 \cap E^c, \ldots$, son Lebesgue medibles. Además, $A \cap E \subset \bigcup_n \left(I_n \cap E\right)$ y $A \cap E^c \subset \bigcup_n \left(I_n \cap E^c\right)$, así que:

$m_e(A \cap E) \leq m_e(\bigcup_n \left(I_n \cap E\right)) \leq \sum_n m_e\left(I_n \cap E\right) = \sum_n m\left(I_n \cap E\right)$,

$m_e(A \cap E^c) \leq m_e(\bigcup_n \left(I_n \cap E^c\right)) \leq \sum_n m_e\left(I_n \cap E^c\right) = \sum_n m\left(I_n \cap E^c\right)$.

Por lo tanto:

$m_e(A \cap E) + m_e(A \cap E^c) \leq \sum_n m\left(I_n \cap E\right) + \sum_n m\left(I_n \cap E^c\right)$

$= \sum_n \left[m\left(I_n \cap E\right) + m\left(I_n \cap E^c\right)\right] = \sum_n m\left(I_n\right) = \sum_n l(I_n)$.

Así que:

$m_e(A \cap E) + m_e(A \cap E^c) \leq \text{ínf}\left\{\sum_n l(I_n) : I_1, I_2, \ldots \text{ son intervalos ajenos y } A \subset \bigcup_n I_n\right\}$

$= m_e(A)$.

Por lo tanto, E satisface la condición de medibilidad de Carathéodory.

Más adelante expondremos de manera detallada la formulación moderna de la teoría de la medida Lebesgue, utilizando como condición de medibilidad la de Carathéodory por ser más fácil de manejar. Además, no será necesario restringirnos a los conjuntos acotados.

1.4. La integral de Lebesgue

La formulación de Riemann del problema de la integral de una función condujo al surgimiento del concepto de contenido y a mostrar cómo se encuentra estrechamente vinculado al de integral, siendo prácticamente dos conceptos equivalentes en el sentido de que con cualquiera de ellos se puede introducir y desarrollar el otro.

Cuando más tarde Borel introdujo el concepto de medida cero y Lebesgue desarrolló una teoría de la medida, más general tanto que la de Jordan como la que había desarrollado Borel, fue posible para el mismo Lebesgue desarrollar una teoría de integración, ahora siguiendo un proceso inverso, es decir, partiendo del concepto de medida para llegar al de integral.

Al igual que la teoría de la medida resultó ser más general que la teoría del contenido, la teoría de la integral desarrollada por Lebesgue resultó ser más general que la teoría de la integral de Riemann.

Es necesario remarcar que Lebesgue desarrolló su teoría de la medida con el objetivo de resolver el problema de la integral que se había planteado. El mismo título del libro que publicó (*Leçons sur l'intégration et la recherche des fonctions primitives*) deja ver claramente que su interés principal era el del concepto de integral. En su libro hizo un estudio del desarrollo del concepto de integral y de las definiciones que diferentes autores habían propuesto, haciendo énfasis en las condiciones para que una función sea integrable. Aclaraba el por qué de su interés de no limitarse al estudio de las funciones para las cuales se puede dar una definición simple de la integral:

> *"si se quisiera limitarse siempre a la consideración de esas buenas funciones, habría que renunciar a la resolución de muchos problemas con enunciados simples planteados desde hace mucho tiempo. Es para la resolución de esos problemas, y no por amor a las complicaciones, que he introducido en este libro una definición de la integral más general que la de Riemann y que la incluye como caso particular."*

Una vez definida la integral, Lebesgue se aboca a estudiar sus propiedades y a utilizarla para profundizar en el estudio de la teoría de funciones: "Como aplicación de la definición de la integral, estudié la búsqueda de funciones primitivas y la rectificación de curvas. A esas dos aplicaciones hubiera querido agregar otra muy importante: el estudio del desarrollo trigonométrico de las funciones; pero en mi curso, no pude dar a ese tema más que indicaciones tan incompletas que he juzgado inútil reproducirlas aquí." Con relación a su definición de integral agregó:

> *“Aquellos que me leerán con empeño, lamentando tal vez que las cosas no sean más simples, pienso que estarán de acuerdo conmigo en que esta definición es necesaria y natural. Me atrevo a decir que es, en un cierto sentido, más simple que la de Riemann, tan fácil de asimilar como ella y que únicamente los hábitos adquiridos anteriormente pueden hacerla parecer más complicada.”*

La definición de Lebesgue de la integral tuvo su motivación directa en la relación que existe entre la integral de Riemann y la teoría del contenido.

Recordando que si $f : [a, b] \to \mathbb{R}$ es una función acotada no negativa y E la región en $\mathbb{R}^2$ acotada por el eje x y la gráfica de f entre a y b, entonces f es Riemann integrable si y sólo si E es Jordan medible y, en ese caso, se tiene $\int_a^b f(x)dx = c(E)$, Lebesgue observó que cuando el conjunto E es (Lebesgue) medible se puede definir la integral de f como $\int_a^b f(x)dx = m(E)$. Automáticamente, esta definición resulta ser una extensión de la integral de Riemann pues si E es Jordan medible también es Lebesgue medible, pero hay conjuntos Lebesgue medibles que no son Jordan medibles. Una vez formulada esta definición geométrica de la integral, Lebesgue se planteó el problema de caracterizar a las funciones integrables y de llegar a la definición de la integral por la vía analítica. El primer problema lo resolvió demostrando el siguiente resultado:

Sea $f : [a, b] \to \mathbb{R}$ una función acotada no negativa, entonces el conjunto

$E = \{(x, y) \in \mathbb{R}^2 : x \in [a, b] \text{ y } y \in [0, f(x)]\}$

es medible si y sólo si el conjunto $\{x \in [a, b] : f(x) > \alpha\}$ es medible para cualquier $\alpha \in \mathbb{R}$.

De aquí que si $f : [a, b] \to \mathbb{R}$ es una función acotada, la defina como medible si el conjunto $\{x \in [a, b] : f(x) > \alpha\}$ es medible para cualquier $\alpha \in \mathbb{R}$.

Se demuestra fácilmente que la suma, el producto y otras operaciones entre funciones medibles resulta ser medible. Demostró Lebesgue una propiedad más cuya importancia resaltaba: el límite de una sucesión convergente de funciones medibles es una función medible.

El segundo problema lo resolvió Lebesgue aproximando la integral de una función medible considerando particiones cada vez más finas del intervalo donde toma valores la función, en lugar de hacerlo, como se hace para definir la integral de Riemann, mediante particiones cada vez más finas del intervalo donde está definida la función. Ésta es una de las ideas originales de Lebesgue en su definición de integral. De manera específica, el razonamiento de Lebesgue es como sigue:

Sea $f : [a, b] \to \mathbb{R}$ una función medible, $\ell = \text{ínf}\{f(x) : x \in [a, b]\}$, $L = \sup\{f(x) : x \in [a, b]\}$ y, dada $\varepsilon > 0$, consideremos una partición del intervalo $[\ell, L]$, $\ell = \ell_0 < \ell_1 < \cdots < \ell_n = L$, de norma menor que ε. Definamos las funciones $\varphi : [a, b] \to \mathbb{R}$ y $\Phi : [a, b] \to \mathbb{R}$ de la siguiente manera:

$\varphi(x) = \sum_{i=0}^{n-1} \ell_i I_{\{y\in[a,b]:\ell_i\leq f(y)<\ell_{i+1}\}}(x) + \ell_n I_{\{y\in[a,b]:f(y)=\ell_n\}}(x)$,

$\Phi(x) = \ell_0 I_{\{y\in[a,b]:f(y)=\ell_0\}}(x) + \sum_{i=0}^{n-1} \ell_{i+1} I_{\{y\in[a,b]:\ell_i<f(y)\leq\ell_{i+1}\}}(x)$.

Para cualquier $x \in [a,b]$, se tiene:

$\varphi(x) \leq f(x) \leq \Phi(x)$,

$\Phi(x) - \varphi(x) = \sum_{i=0}^{n-1} (\ell_{i+1} - \ell_i) I_{\{y\in[a,b]:\ell_i<f(y)<\ell_{i+1}\}}(x) < \varepsilon \sum_{i=0}^{n-1} I_{\{y\in[a,b]:\ell_i<f(y)<\ell_{i+1}\}}(x) \leq \varepsilon$.

Denotemos por λ a la medida de Lebesgue en el intervalo $[a,b]$ y definamos:

$\sigma = \sum_{i=0}^{n-1} \ell_i \lambda(\{y \in [a,b] : \ell_i \leq f(y) < \ell_{i+1}\}) + \ell_n \lambda(\{y \in [a,b] : f(y) = \ell_n\})$

$= \sum_{i=0}^{n-1} \ell_i \lambda(\{y \in [a,b] : \ell_i < f(y) < \ell_{i+1}\}) + \sum_{i=0}^{n} \ell_i \lambda(\{y \in [a,b] : f(y) = \ell_i\})$,

$\Sigma = \ell_0 \lambda(\{y \in [a,b] : f(y) = \ell_0\}) + \sum_{i=0}^{n-1} \ell_{i+1} \lambda(\{y \in [a,b] : \ell_i < f(y) \leq \ell_{i+1}\})$

$= \sum_{i=0}^{n-1} \ell_{i+1} \lambda(\{y \in [a,b] : \ell_i < f(y) < \ell_{i+1}\}) + \sum_{i=0}^{n} \ell_i \lambda(\{y \in [a,b] : f(y) = \ell_i\})$.

Observemos que se tiene:

$0 \leq \Sigma - \sigma = \sum_{i=0}^{n-1} (\ell_{i+1} - \ell_i) \lambda(\{y \in [a,b] : \ell_i < f(y) < \ell_{i+1}\})$

$< \varepsilon \sum_{i=0}^{n-1} \lambda(\{y \in [a,b] : \ell_i < f(y) < \ell_{i+1}\}) \leq \varepsilon(b-a)$.

Tomemos ahora una sucesión decreciente $(\varepsilon_m)_{m\in\mathbb{N}}$ que converja a cero.

Para cada $m \in \mathbb{N}$, definamos las funciones $\varphi_m : [a,b] \to \mathbb{R}$ y $\Phi_m : [a,b] \to \mathbb{R}$, así como los números reales σ_m y Σ_m, como antes, tomando $\varepsilon = \varepsilon_m$. Tomemos las particiones del intervalo $[\ell, L]$ de tal forma que, para cualquier $m \in \mathbb{N}$, la partición $\left\{\ell_0^{(m+1)}, \ell_1^{(m+1)}, \cdots < \ell_{n_{m+1}}^{(m+1)}\right\}$, correspondiente a ε_{m+1}, es un refinamiento de la partición $\left\{\ell_0^{(m)}, \ell_1^{(m)}, \cdots < \ell_{n_m}^{(m)}\right\}$ correspondiente a ε_m. Se tiene entonces que la sucesión de funciones $(\varphi_m)_{m\in\mathbb{N}}$ es creciente, mientras que la sucesión $(\Phi_m)_{m\in\mathbb{N}}$ es decreciente, así que ambas convergen puntualmente. También, la sucesión $(\sigma_m)_{m\in\mathbb{N}}$ es creciente y la sucesión $(\Sigma_m)_{m\in\mathbb{N}}$ es decreciente, así que ambas son convergentes.

Además, como $\varphi_m(x) \leq f(x) \leq \Phi_m(x)$ y $\Phi_m(x) - \varphi_m(x) \leq \varepsilon_m$ para cualquier $m \in \mathbb{N}$ y cualquier $x \in [a,b]$, las sucesiones $(\varphi_m)_{m\in\mathbb{N}}$ y $(\Phi_m)_{m\in\mathbb{N}}$ convergen uniformemente a la función f.

También, como $0 \leq \Sigma_m - \sigma_m < \varepsilon_m(b-a)$, las sucesiones $(\sigma_m)_{m\in\mathbb{N}}$ y $(\Sigma_m)_{m\in\mathbb{N}}$ convergen al mismo valor. La integral $\int_a^b f(x)\,dx$ se define como este límite común.

Falta demostrar que se obtiene el mismo valor de la integral para cualquier sucesión decreciente $(\varepsilon_m)_{m\in\mathbb{N}}$ que converja a cero, lo cual hace Lebesgue de la siguiente manera:

Consideremos otra sucesión $(\varepsilon'_m)_{m\in\mathbb{N}}$ que converja a cero.

Para la sucesión $(\varepsilon_m)_{m\in\mathbb{N}}$ inicial, tenemos una sucesión de particiones $(P_m)_{m\in\mathbb{N}}$ del intervalo $[\ell, L]$, y para la sucesión $(\varepsilon'_m)_{m\in\mathbb{N}}$, tenemos una sucesión de particiones $(P'_m)_{m\in\mathbb{N}}$ del intervalo $[\ell, L]$. Para cada $m \in \mathbb{N}$, definamos:

$P''_m = P_m \cup P'_m$.

Se tiene entonces, para cualquier $m \in \mathbb{N}$:

$\sigma_m \leq \sigma''_m \leq \Sigma''_m \leq \Sigma_m$,

$\sigma'_m \leq \sigma''_m \leq \Sigma''_m \leq \Sigma'_m$.

De la primeras desigualdades, se sigue que las sucesiones $(\sigma''_m)_{m\in\mathbb{N}}$ y $(\Sigma''_m)_{m\in\mathbb{N}}$ convergen al mismo valor que las sucesiones $(\sigma_m)_{m\in\mathbb{N}}$ y $(\Sigma_m)_{m\in\mathbb{N}}$.

De las segundas desigualdades, se sigue que las sucesiones $(\sigma''_m)_{m\in\mathbb{N}}$ y $(\Sigma''_m)_{m\in\mathbb{N}}$ convergen al mismo valor que las sucesiones $(\sigma'_m)_{m\in\mathbb{N}}$ y $(\Sigma'_m)_{m\in\mathbb{N}}$.

Así que, las sucesiones $(\sigma_m)_{m\in\mathbb{N}}$ y $(\Sigma_m)_{m\in\mathbb{N}}$ convergen al mismo valor que las sucesiones $(\sigma'_m)_{m\in\mathbb{N}}$ y $(\Sigma'_m)_{m\in\mathbb{N}}$.

Obsérvese que si $E \subset [a, b]$, entonces $\int_a^b I_E(x)\, dx = \lambda(E)$.

Una vez definida la integral de una función medible, se demuestran fácilmente las 5 primeras propiedades que Lebesgue planteó como propiedades que debe tener la integral.

La sexta propiedad es un corolario del siguiente resultado que demostró Lebesgue, el cual es ahora conocido como el teorema de la convergencia uniformemente acotada:

Si una sucesión de funciones medibles $(f_n)_{n\in\mathbb{N}}$, definidas sobre un intervalo $[a, b]$, converge a una función f y existe $M \in \mathbb{R}$ tal que $|f_n(x)| \leq M$ para cualquier $x \in [a, b]$ y cualquier $n \in \mathbb{N}$, entonces:

$$\text{lím}_{n\rightsquigarrow\infty} \int_a^b f_n(x)\, dx = \int_a^b f(x)\, dx.$$

1.5. La medida de Lebesgue en $\mathbb{R}^2$

Como puede verse, en la caracterización de las funciones integrables, Lebesgue hacía referencia a conjuntos medibles en $\mathbb{R}^2$. En su libro, Lebesgue planteó que éstos se pueden definir de manera similar a como lo hacía en el caso de los conjuntos medibles en $\mathbb{R}$. Para ello se requiere extender a $\mathbb{R}^2$ el resultado de Borel que asegura que si un intervalo cerrado y acotado es cubierto por una infinidad numerable de intervalos abiertos, entonces existe una colección finita de esos intervalos que también lo cubren y, una vez hecho esto, extender también la propiedad básica que permite definir la medibilidad con el método de Lebesgue, a saber, que si I es un intervalo cerrado y acotado y $(I_j)_{j\in\mathbb{N}}$ una sucesión de intervalos abiertos tales que $I \subset \bigcup_{j=1}^{\infty} I_j$, entonces $l(I) \leq \sum_{j=1}^{\infty} l(I_j)$. El resultado de Borel lo demostró Lebesgue utilizando un resultado de Peano que afirma que se puede cubrir todo $\mathbb{R}^2$ mediante una curva continua. La demostración ahora clásica del resultado de Borel es como sigue:

PROPOSICIÓN 1.4. *Sea $R = [a, b] \times [c, d]$ un rectángulo acotado en $\mathbb{R}^2$ y una sucesión $R^{(j)} = \left(a_1^{(j)}, b_1^{(j)}\right) \times \left(a_2^{(j)}, b_2^{(j)}\right)$ (con $j \in \mathbb{N}$) de rectángulos no vacíos en $\mathbb{R}^2$ tales que $R \subset \bigcup_{j=1}^{\infty} R^{(j)}$, entonces existe un número finito de rectángulos de la familia $\left\{R^{(j)} : j \in \mathbb{N}\right\}$ cuya unión contiene a R.*

Demostración

Supongamos que no existe un número finito de rectángulos de la familia $\left\{R^{(j)} : j \in \mathbb{N}\right\}$ cuya unión contenga a R.

Denotemos por c_1 al punto medio del intervalo $[a_1, b_1]$ y por c_2 al punto medio del intervalo $[a_2, b_2]$.

Consideremos los rectángulos $[a_1, c_1] \times [a_2, c_2]$, $[a_1, c_1] \times [c_2, b_2]$, $[c_1, b_1] \times [a_2, c_2]$ y $[c_1, b_1] \times [c_2, b_2]$, cuya unión es R. Como no existe un número finito de rectángulos de la familia $\left\{R^{(j)} : j \in \mathbb{N}\right\}$ cuya unión contenga a R, por lo menos uno de esos 4 rectángulos tiene la misma propiedad. Denotemos por R_1 a cualquiera de los 4 rectángulos para el cual no existe un número finito de rectángulos de la familia $\left\{R^{(j)} : j \in \mathbb{N}\right\}$ cuya unión lo contenga.

Sea $R_1 = \left[a^{(1)}, b^{(1)}\right] \times \left[c^{(1)}, d^{(1)}\right]$.

Repitamos el procedimiento, partiendo ahora R_1 en cuatro rectángulos y denotemos por R_2 a cualquiera de los 4 rectángulos para el cual no existe un número finito de rectángulos de la familia $\left\{R^{(j)} : j \in \mathbb{N}\right\}$ cuya unión lo contenga.

Sea $R_2 = \left[a^{(2)}, b^{(2)}\right] \times \left[c^{(2)}, d^{(2)}\right]$.

Continuando con este procedimiento, obtenemos una sucesión de rectangulos encajados $(R_k)_{k\in\mathbb{N}}$ para cada uno de los cuales no existe un número finito de rectángulos de la familia $\left\{R^{(j)} : j \in \mathbb{N}\right\}$ cuya unión lo contenga.

Denotemos por L a la longitud de la diagonal del rectángulo R y por L_k a la longitud de la diagonal del rectángulo R_k. Por la manera en que los construimos, se tiene:

$L_k = \frac{1}{2^k} L$.

Como los rectángulos R_k son cerrados y están encajados, su intersección es no vacía. Además, como la longitud de su diagonal tiende a cero, la intersección es un único punto $x_0 = (a_0, b_0)$, el cual pertenece a R ya que R es cerrado.

Sea $R^{(j_0)}$ un intervalo de la familia $\left\{R^{(j)} : j \in \mathbb{N}\right\}$ al cual pertenezca x_0 y sea $B_r(x_0)$ una bola abierta, de radio r y centro x_0, contenida en $R^{(j_0)}$. Tomemos $k_0 \in \mathbb{N}$ tal que $L_{k_0} < r$;

entonces, como $x \in R_{k_0}$, la distancia de cualquier punto de R_{k_0} a x_0 es menor que r. Por lo tanto, $R_{k_0} \subset B_r(x_0) \subset R^{(j_0)}$, lo cual es una contradicción.

■

Con este resultado la otra parte ya es simple:

DEFINICIÓN 1.5. *Si $R = I_1 \times I_2$ es un rectángulo en $\mathbb{R}^2$, denotaremos por $a(R)$ al área de R, es decir al producto $l(I_1)\, l(I_2)$, donde $l(I)$ es la longitud del intervalo I.*

LEMA 1.1. *Sea $R = [a_1, b_1] \times [a_2, b_2]$ un rectángulo acotado en $\mathbb{R}^2$ y una colección finita de rectángulos abiertos no vacíos $\left\{R^{(j)} = \left(a_1^{(j)}, b_1^{(j)}\right) \times \left(a_2^{(j)}, b_2^{(j)}\right) : \ m \in \mathbb{N} \text{ y } j \in \{1, 2, \ldots, m\}\right\}$ tales que $R \subset \bigcup_{j=1}^m R^{(j)}$, entonces:*

$a(R) \leq \sum_{j=1}^m a\left(R^{(j)}\right)$.

Demostración

Si alguno de los rectángulos $R^{(j)}$ no está acotado, el resultado es inmediato, así que asumiremos que todos los rectángulo $R^{(j)}$ están acotados. Para cada $j \in \{1, 2, \ldots, m\}$, denotemos por $\overline{R}^{(j)}$ a la cerradura de $R^{(j)}$, es decir al rectángulo $\left[a_1^{(j)}, b_1^{(j)}\right] \times \left[a_2^{(j)}, b_2^{(j)}\right]$.

Para cada $i \in \{1, 2\}$, los puntos $a_i, b_i, a_i^{(1)}, b_i^{(1)}, \ldots, a_i^{(m)}, b_i^{(m)}$ constituyen una partición de un intervalo $[c_i, d_i]$. Para $i = 1$ (resp. $i = 2$), denotemos por $x_0^{(1)}, x_1^{(1)}, x_2^{(1)}, \ldots, x_{n_1}^{(1)}$ (resp. $x_0^{(2)}, x_1^{(2)}, x_2^{(2)}, \ldots, x_{n_2}^{(2)}$) los elementos de esa partición, ordenados del menor al mayor y, para $k_1 \in \{1, 2, \ldots, n_1\}$ y $k_2 \in \{1, 2, \ldots, n_2\}$, definamos $R_{k_1,k_2} = \left(x_{k_1-1}^{(1)}, x_{k_1}^{(1)}\right) \times \left(x_{k_2-1}^{(2)}, x_{k_2}^{(2)}\right)$ y denotemos por $\overline{R}_{k_1,k_2}$ a la cerradura de R_{k_1,k_2}, es decir al rectángulo $\left[x_{k_1-1}^{(1)}, x_{k_1}^{(1)}\right] \times \left[x_{k_2-1}^{(2)}, x_{k_2}^{(2)}\right]$.

Por la construcción de los rectángulos R_{k_1,k_2}, el rectángulo R, así como cada uno de los rectángulos $\overline{R}^{(j)}$, con $j \in \{1, 2, \ldots, m\}$, es la unión de algunos de los rectángulos $\overline{R}_{k_1,k_2}$. Además, cualquier par de rectángulos R_{k_1,k_2} son ajenos.

Sean $R_1, R_2, \ldots, R_t$ los rectángulos de la familia

$\left\{\overline{R}_{k_1,k_2} : k_1 \in \{1, 2, \ldots, n_1\} \text{ y } k_2 \in \{1, 2, \ldots, n_2\}\right\}$,

cuya unión es igual a R.

Como $R \subset \bigcup_{j=1}^m R^{(j)}$, cada rectángulo R_i ($i \in \{1, 2, \ldots, t\}$) está contenido en algún rectángulo $\overline{R}^{(j)}$ ($j \in \{1, 2, \ldots, m\}$). Además:

$a(R) = \sum_{i=1}^t a(R_i)$.

Para cada $j \in \{1, 2, \ldots, m\}$, sea D_j el conjunto de índices de los rectángulos de la familia $\{R_1, R_2, \ldots, R_t\}$ que están contenidos en $\overline{R}^{(j)}$. Obviamente se tiene:

$a\left(R^{(j)}\right) = a\left(\overline{R}^{(j)}\right) \geq \sum_{\{i \in D_j\}} a\left(R_i\right)$.

Y como cada rectángulo R_i ($i \in \{1, 2, \ldots, t\}$) está contenido en algún rectángulo $\overline{R}^{(j)}$ ($j \in \{1, 2, \ldots, m\}$), se tiene:

$\sum_{j=1}^{m} a\left(R^{(j)}\right) \geq \sum_{j=1}^{m} \sum_{\{i \in D_j\}} a\left(R_i\right) \geq \sum_{i=1}^{t} a\left(R_i\right) = a\left(R\right)$.

■

PROPOSICIÓN 1.5. *Sea* $R = [a_1, b_1] \times [a_2, b_2]$ *un rectángulo acotado en* $\mathbb{R}^2$ *y una sucesión* $R^{(j)} = \left(a_1^{(j)}, b_1^{(j)}\right) \times \left(a_2^{(j)}, b_2^{(j)}\right)$ *(con* $j \in \mathbb{N}$*) de rectángulos no vacíos en* $\mathbb{R}^2$ *tales que* $R \subset \bigcup_{j=1}^{\infty} R^{(j)}$*, entonces:*

$a\left(R\right) \leq \sum_{j=1}^{\infty} a\left(R^{(j)}\right)$.

Demostración

Por el teorema de Borel, existe una colección finita, $R^{(j_1)}, R^{(j_2)}, \ldots, R^{(j_m)}$, tal que:

$R \subset \bigcup_{i=1}^{m} R^{(j_1)}$.

Así que, por el lema anterior:

$a\left(R\right) \leq \sum_{i=1}^{m} a\left(R^{(j_i)}\right) \leq \sum_{j=1}^{\infty} a\left(R^{(j)}\right)$.

■

PROPOSICIÓN 1.6. *Sea* R *un rectángulo finito de cualquier tipo y* $R_1, R_2, \ldots$ *una familia finita o infinita numerable de rectángulos abiertos tales que* $R \subset \bigcup_j R_j$*, entonces:*

$a(R) \leq \sum_j a(R_j)$.

Demostración

Sea $R = I \times J$, a y b los extremos del intervalo I, c y d los extremos del intervalo J. Definamos:

$L_1 = \{(a, y) : y \in [c, d]\}$,

$L_2 = \{(b, y) : y \in [c, d]\}$,

$L_3 = \{(x, c) : x \in [a, b]\}$,

$L_4 = \{(x, d) : x \in [a, b]\}$.

Dada $\varepsilon > 0$, para $k \in \{1, 2, 3, 4\}$, sea $R^{(k)}$ un rectángulo abiertos que contengan a L_k y de área igual a $\frac{\varepsilon}{4}$. Entonces la unión de los rectángulos $R^{(1)}, R^{(2)}, R^{(3)}, R^{(4)}, R_1, R_2, \ldots$ contiene al rectángulo $\overline{R} = [a, b] \times [c, d]$. Por el teorema de Borel, existe entonces una colección finita de esos rectángulos cuya unión contiene a $\overline{R}$, así que:

$a(R) = a(\overline{R}) \leq \sum_j a(R_j) + \varepsilon.$

Como esta relación es válida para cualquier $\varepsilon > 0$, se puede concluir que $a(R) \leq \sum_j a(R_j)$.

■

Con base en lo anterior, la definición de los conjuntos medibles en $\mathbb{R}^2$ y la demostración de sus propiedades, puede hacerse siguiendo paso a paso el razonamiento de Lebesgue para el caso de los conjuntos medibles en $\mathbb{R}$.

CAPÍTULO 2

MEDIDA DE LEBESGUE

2.1. Álgebras, σ-álgebras y borelianos

Una vez desarrollada la teoría iniciada por Borel y Lebesgue, se volvieron básicos los conceptos de álgebra y σ-álgebra de subconjuntos de un conjunto dado, ya que la familia de conjuntos para los cuales se pudo definir una medida, si bien no siempre está formada por todos los subconjuntos del conjunto, constituye una σ-álgebra.

DEFINICIÓN 2.1 (**álgebra**). *Sea $\mathbb{F}$ un conjunto. Se dice que una familia $\mathcal{A}$ de subconjuntos de $\mathbb{F}$ es un álgebra si se satisfacen las siguientes condiciones:*

(i) $\mathbb{F} \in \mathcal{A}$.
(ii) *Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$.*
(iii) *Si $A_1, \ldots, A_n$ es cualquier familia finita de elementos de $\mathcal{A}$, entonces $\bigcup_{k=1}^{n} A_k \in \mathcal{A}$.*

DEFINICIÓN 2.2 (**$\boldsymbol{\sigma}$-álgebra**). *Sea $\mathbb{F}$ un conjunto. Se dice que una familia $\Im$ de subconjuntos de $\mathbb{F}$ es una σ-álgebra si es un álgebra y dada cualquier colección infinita numerable de elementos de $\Im$, $A_1, A_2, \ldots$, entonces $\bigcup_{k=1}^{\infty} A_k \in \Im$.*

DEFINICIÓN 2.3. *Llamaremos espacio medible a una pareja $(\mathbb{F}, \Im)$, donde $\mathbb{F}$ es un conjunto y $\Im$ una σ-álgebra de subconjuntos de $\mathbb{F}$.*

Si $\mathbb{F}$ es un conjunto arbitrario, se pueden definir distintas σ-álgebras de subconjuntos de $\mathbb{F}$. Por ejemplo, la familia $\Im = \{\emptyset, \mathbb{F}\}$ constituye una σ-álgebra de subconjuntos de $\mathbb{F}$. También la familia formada por todos los subconjuntos de $\mathbb{F}$ constituye una σ-álgebra de subconjuntos de $\mathbb{F}$. Si $A_1, A_2, \ldots, A_n$ son subconjuntos de $\mathbb{F}$, $T = \{1, 2, \ldots, n\}$, $U = \{i_1, \ldots, i_k\} \subset T$ y $T - U = \{j_1, \ldots, j_{n-k}\}$, definamos:

$B_U = A_{i_1} \cap \cdots \cap A_{i_k} \cap A^c_{j_1} \cap \cdots \cap A^c_{j_{m-k}}$,

$\mathcal{U} = \{B_U : U \subset T\}$.

Entonces, la familia $\mathcal{A}$ formada por el conjunto vacío y las uniones de elementos de $\mathcal{U}$ constituye una σ-álgebra de subconjuntos de $\mathbb{F}$.

Los anteriores son ejemplos sencillos y, podría decirse, artificiales. En general vamos a trabajar con conjuntos $\mathbb{F}$ que son infinitos y con σ-álgebras que están formadas por una infinidad de subconjuntos de $\mathbb{F}$.

La necesidad de introducir el concepto de σ-álgebra proviene de que, en general, para definir una medida se sigue un procedimiento similar al que siguió Lebesgue para definir lo que podríamos llamar la longitud de un subconjunto de los números reales. Partió de que la longitud de un intervalo de extremos a y b está definida como la diferencia $b-a$ y se planteó entonces el problema de extender el concepto de longitud a todos los subconjuntos de los números reales. A lo que llegó es que es posible realizar esa extensión hasta abarcar una determinada familia de subconjuntos. Mostró también que la familia de conjuntos hasta donde es posible llevar su proceso de extensión, si bien no necesariamente está formada por todos los subconjuntos de $\mathbb{R}$, es una familia bastante grande ya que es cerrada bajo complementos y uniones e interserciones numerables; es decir, constituye lo que estamos definiendo como σ-álgebra. En general, el procedimiento de Lebesgue es el que se sigue para definir una medida: se comienza por asignar una medida a cada conjunto de una determinada familia y después se realiza el proceso de extensión. Con este camino siempre se llega a definir la medida sobre una familia de conjuntos que forma una σ-álgebra.

Definición 2.4 (**Intersección de $\boldsymbol{\sigma}$-álgebras**). *Dado un conjunto $\mathbb{F}$ y una familia arbitraria de σ-álgebras de subconjuntos de $\mathbb{F}$, se define la intersección de esas σ-álgebras como la familia de conjuntos que pertenecen a todas ellas.*

Se puede ver fácilmente que la intersección de σ-álgebras de subconjuntos de $\mathbb{F}$ es también una σ-álgebra y, dada una colección arbitraria $\mathcal{B}$ de subconjuntos de $\mathbb{F}$, siempre existe por lo menos una σ-álgebra que contiene a todos los elementos de $\mathcal{B}$, a saber, la formada por todos los subconjuntos de $\mathbb{F}$. Se puede definir entonces una σ-álgebra como la intersección de todas las σ-álgebras de subconjuntos de $\mathbb{F}$ que contienen a todos los elementos de $\mathcal{B}$.

Definición 2.5 (**$\boldsymbol{\sigma}$ álgebra generada por una familia de conjuntos**). *Dada una colección $\mathcal{A}$ de subconjuntos de un conjunto $\mathbb{F}$, se define la σ-álgebra generada por $\mathcal{A}$ como la intersección de todas las σ-álgebras que contienen a todos los conjuntos de $\mathcal{A}$. Denotaremos por $\sigma(\mathcal{A})$ a esta σ-álgebra.*

Evidentemente la σ-álgebra generada por $\mathcal{A}$ es la más pequeña σ-álgebra de subconjuntos de $\mathbb{F}$ que contiene a todos los elementos de $\mathcal{A}$.

Una σ-álgebra de particular importancia es la de los conjuntos borelianos en $\mathbb{R}$ y, de manera más general, en $\mathbb{R}^n$.

Los conjuntos borelianos deben su nombre a Émile Borel quien los introdujo para caracterizar a los subconjuntos de $\mathbb{R}$ a los cuales se les puede asignar una longitud.

La idea fundamental consiste en que se puede asignar una longitud a todos los subconjuntos de $\mathbb{R}$ que se puedan obtener a partir de los intervalos mediante las operaciones conjuntistas de unión numerable y diferencia. La definición moderna se basa en la generación de σ-álgebras de acuerdo con la definición anterior.

Cabe mencionar que Borel era constructivista, de manera que únicamente aceptaba definiciones de elementos para los cuales se pudiera decir, explicitamente, como se construían. La siguiente definición de los conjuntos borelianos, obviamente, no es la de Borel, ya que no se dice en ella cómo se construye cada conjunto boreliano.

DEFINICIÓN 2.6 (**$\boldsymbol{\sigma}$-álgebra de Borel en** $\mathbb{R}$). *La σ-álgebra de Borel en $\mathbb{R}$, la cual será denotada por $\mathfrak{B}(\mathbb{R})$, es la σ-álgebra de subconjuntos de $\mathbb{R}$ generada por la familia de todos los intervalos de números reales. A los elementos de esa σ-álgebra los llamaremos borelianos de $\mathbb{R}$.*

Las siguientes propiedades pueden ser probada fácilmente, para ello sólo hay que demostrar que, a partir de cada una de las familias que se dan, se puede obtener un intervalo de cualquier tipo utilizando las operaciones de complemento y de uniones o intersecciones numerables.

EJERCICIO 2.1. *Demuestra que la σ-álgebra de los conjuntos borelianos de $\mathbb{R}$ está generada por cualquiera de las siguientes familias de conjuntos.*

(i) *Los intervalos de la forma $(-\infty, x]$, donde $x \in \mathbb{R}$.*
(ii) *Los intervalos de la forma $(-\infty, x)$, donde $x \in \mathbb{R}$.*
(iii) *Los intervalos de la forma $(a, b]$, donde $a, b \in \mathbb{R}$.*
(iv) *Los intervalos de la forma $[a, b)$, donde $a, b \in \mathbb{R}$.*
(v) *Los intervalos de la forma (a, b), donde $a, b \in \mathbb{R}$.*
(vi) *Los intervalos de la forma $[a, b]$, donde $a, b \in \mathbb{R}$.*

EJERCICIO 2.2. *Muestra que no todo boreliano de $\mathbb{R}$ es una unión numerable de intervalos o el complemento de una unión numerable de intervalos.*

Sugerencia: Denotemos por C al conjunto de números racionales contenidos en el intervalo $(-\infty, 0)$ y por D al conjunto de números racionales contenidos en el intervalo $[0, \infty)$.

Definamos:

$A = ((-\infty, 0) - C) \cup D$

Demuestra lo siguiente:

1. A es un conjunto boreliano de $\mathbb{R}$.

2. $A \cap (-\infty, 0)$ y $A^c \cap [0, \infty)$ son conjuntos no numerables que no contienen ningún intervalo de longitud positiva.

3. A no es una unión numerable de intervalos ni el complemento de una unión numerable de intervalos.

Definición 2.7. *Cuando los extremos de un intervalo sean números reales, diremos que el intervalo es finito (esto equivale a decir que el intervalo es un conjunto acotado).*

Conviene considerar desde este momento al conjunto de números reales extendidos, el cual consiste del conjunto de números reales y dos elementos especiales, $-\infty$ y ∞, con los cuales operaremos bajo las siguientes convenciones:

Si $c \in \mathbb{R}$, entonces:

$-\infty < c < \infty$,

$c - \infty = -\infty$,

$c + \infty = \infty$,

$c(\infty) = \infty$ si $c > 0$,

$c(\infty) = -\infty$ si $c < 0$,

$(0)(\infty) = (0)(-\infty) = 0$,

$\frac{c}{\infty} = \frac{c}{-\infty} = 0$,

$(\infty)(\infty) = \infty + \infty = \infty$,

$\infty - \infty$ e $\frac{\infty}{\infty}$ no están definidos.

$\overline{\mathbb{R}}$ denotará al conjunto $\mathbb{R} \cup \{-\infty, \infty\}$.

Tomando en consideración las convenciones anteriores, las propiedades usuales de conmutatividad y asociatividad de las operaciones de suma y producto entre números reales siguen siendo válidas, cuando estén definidas, sobre $\overline{\mathbb{R}}$.

El conjunto de números reales $\mathbb{R}$ también será denotado por $(-\infty, \infty)$. El conjunto de números reales no negativos será denotado por $[0, \infty)$, o por $\mathbb{R}^+$, y $\overline{\mathbb{R}}^+$ denotará al conjunto $\mathbb{R}^+ \cup \{\infty\}$.

$\overline{\mathbb{R}}^n$ denotará al conjunto $\{(x_1, \ldots, x_n) : x_k \in \overline{\mathbb{R}}$ para cualquier $k \in \{1, 2, \ldots, n\}\}$. Cuando estén bien definidas, consideraremos sobre $\overline{\mathbb{R}}^n$ las operaciones usuales definidas en $\mathbb{R}^n$.

Denotaremos por $\overline{\mathcal{R}}$ a la familia de subconjuntos de $\overline{\mathbb{R}}^n$ de la forma $I_1 \times I_2 \times \cdots \times I_n$, donde, para cualquier $k \in \{1, 2, \ldots, n\}$, I_k es ya sea un intervalo de la forma $(-\infty, x]$, con $x \in \overline{\mathbb{R}}$, o bien $I_k = \overline{\mathbb{R}}$.

Si $\mathbb{F}$ es un conjunto cualquiera y f y g son dos funciones con valores en $\overline{\mathbb{R}}$, definidas sobre $\mathbb{F}$, vamos a considerar a la suma de f y g como la función $h : \mathbb{F} \to \overline{\mathbb{R}}$ definida de la siguiente manera:

$$h(x) = \begin{cases} f(x) + g(x) & \text{si } f(x) + g(x) \text{ está definida} \\ 1 & \text{si } f(x) + g(x) \text{ no está definida} \end{cases}$$

Esta convención se traslada a la resta de dos funciones ya que $f - g = f + (-g)$.

Definición 2.8 (**σ-álgebra de Borel en $\overline{\mathbb{R}}$**). *La σ-álgebra de Borel en $\overline{\mathbb{R}}$, la cual será denotada por $\mathfrak{B}\left(\overline{\mathbb{R}}\right)$, es la σ-álgebra de subconjuntos de $\overline{\mathbb{R}}$ generada por la familia de intervalos de la forma $(-\infty, x]$, donde $x \in \overline{\mathbb{R}}$. A los elementos de esa σ-álgebra los llamaremos borelianos de $\overline{\mathbb{R}}$.*

Ejercicio 2.3. *Los conjuntos $\{\infty\}$, $\{-\infty\}$ y $\{-\infty, \infty\}$ son borelianos de $\overline{\mathbb{R}}$.*

Proposición 2.1. *La σ-álgebra de Borel en $\overline{\mathbb{R}}$ está formada los borelianos de $\mathbb{R}$ y los conjuntos de la forma $B \cup \{\infty\}$, $B \cup \{-\infty\}$ y $B \cup \{-\infty, \infty\}$, donde B es un boreliano de $\mathbb{R}$.*

Demostración

Sea $\mathcal{H}$ la familia de conjuntos formada los borelianos de $\mathbb{R}$ y los conjuntos de la forma $B \cup \{\infty\}$, $B \cup \{-\infty\}$ y $B \cup \{-\infty, \infty\}$, donde B es un boreliano de $\mathbb{R}$.

Todos los elementos de $\mathcal{H}$ son borelianos de $\overline{\mathbb{R}}$ y $\mathcal{H}$ es una σ-álgebra que contiene a todos los intervalos de la forma $(-\infty, x]$, donde $x \in \overline{\mathbb{R}}$.

■

Proposición 2.2. *La σ-álgebra de Borel en $\overline{\mathbb{R}}$ está generada por la familia de intervalos de la forma $[-\infty, x]$, donde $x \in \mathbb{R}$.*

Demostración

Para cualquier $x \in \mathbb{R}$, se tiene:

$[-\infty, x] = (-\infty, x] \cup \{-\infty\} \in \mathfrak{B}\left(\overline{\mathbb{R}}\right)$.

Sea J la familia de intervalos en $\overline{\mathbb{R}}$ de la forma $[-\infty, x]$, donde $x \in \mathbb{R}$. Entonces:

$\{-\infty\} = \bigcap_{n=1}^{\infty} [-\infty, -n] \in \sigma(J)$,

$(-\infty, \infty] = \overline{\mathbb{R}} - \{-\infty\} \in \sigma(J)$.

Finalmente, si $x \in \mathbb{R}$, entonces:

$(-\infty, x] = [-\infty, x] - \{-\infty\} \in \sigma(J)$.

■

Corolario 2.1. *La σ-álgebra de Borel en $\overline{\mathbb{R}}$ está generada por la familia de intervalos de la forma $[-\infty, x]$, donde $x \in \overline{\mathbb{R}}$.*

Proposición 2.3. *La σ-álgebra de Borel en $\overline{\mathbb{R}}$ está generada por la familia de intervalos de la forma $[-\infty, x)$, donde $x \in \mathbb{R}$.*

Demostración

Sea I la familia de intervalos en $\overline{\mathbb{R}}$ de la forma $[-\infty, x)$, donde $x \in \mathbb{R}$, y J la familia de intervalos en $\overline{\mathbb{R}}$ de la forma $[-\infty, x]$, donde $x \in \mathbb{R}$. Entonces, para cualquier $x \in \mathbb{R}$, se tiene:

$[-\infty, x) = \bigcup_{n=1}^{\infty} \left[-\infty, x - \frac{1}{n}\right] \in \sigma(J)$,

$[-\infty, x] = \bigcap_{n=1}^{\infty} \left[-\infty, x + \frac{1}{n}\right) \in \sigma(I)$.

Así que $\sigma(I) = \sigma(J) = \mathfrak{B}\left(\overline{\mathbb{R}}\right)$.

■

2.2. σ-álgebra de Borel en $\mathbb{R}^n$

DEFINICIÓN 2.9. *Por una celda en $\mathbb{R}^n$ se entenderá un conjunto de la forma $I_1 \times \cdots \times I_n$, donde $I_1, \cdots, I_n$ son intervalos en $\mathbb{R}$.*

Obviamente, si $R = I_1 \times \cdots \times I_n$ entonces R es un conjunto acotado si y sólo si los intervalos $I_1, \cdots, I_n$ son finitos. De la misma manera, R es un conjunto abierto (resp. cerrado) si y sólo si los intervalos $I_1, \cdots, I_n$ son abiertos (resp. cerrados).

Denotaremos por $\mathcal{R}$ a la familia de celdas en $\mathbb{R}^n$.

DEFINICIÓN 2.10 (**$\boldsymbol{\sigma}$-álgebra de Borel en $\mathbb{R}^n$**). *La σ-álgebra de Borel en $\mathbb{R}^n$, la cual será denotada por $\mathfrak{B}(\mathbb{R}^n)$, es la σ-álgebra de subconjuntos de $\mathbb{R}^n$ generada por $\mathcal{R}$. A los elementos de esa σ-álgebra los llamaremos borelianos de $\mathbb{R}^n$.*

PROPOSICIÓN 2.4. *La σ-álgebra generada por la familia de celdas de $\mathbb{R}^n$ de la forma $(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]$, donde $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, contiene a todos los subconjuntos de $\mathbb{R}^n$ de la forma $B_1 \times B_2 \times \ldots \times B_n$, donde $B_1, \ldots, B_n$ son borelianos de $\mathbb{R}$.*

Demostración

Sea $\mathcal{H}$ la σ-álgebra de subconjuntos de $\mathbb{R}^n$ generada por la familia de celdas de $\mathbb{R}^n$ de la forma $(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]$, donde $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

Sea $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ y $U \subset \{1, 2, \ldots, n\}$. Para cada $k \in \{1, 2, \ldots, n\}$, definamos la sucesión de intervalos $\left(J_k^{(m)}\right)_{m \in \mathbb{N}}$ y el intervalos J_k de la siguiente manera:

$$J_k = \begin{cases} (-\infty, x_k] & \text{si } k \notin U \\ \mathbb{R} & \text{si } k \in U \end{cases}$$

$$J_k^{(m)} = \begin{cases} (-\infty, x_k] & \text{si } k \notin U \\ (-\infty, m] & \text{si } k \in U \end{cases}$$

Se tiene:

$J_1 \times J_2 \times \cdots \times J_n = \bigcup_{m=1}^{\infty} J_1^{(m)} \times J_2^{(m)} \times \cdots \times J_n^{(m)}$.

Así que $J_1 \times J_2 \times \cdots \times J_n \in \mathcal{H}$.

Definamos $\mathcal{G} = \{(-\infty, x] : x \in \mathbb{R}\} \cup \{\mathbb{R}\}$.

Por lo anterior, $I_1 \times \cdots \times I_n \in \mathcal{H}$, para cualquier celda $I_1 \times \cdots \times I_n$ donde $I_k \in \mathcal{G}$ para cualquier $k \in \{1, 2, \ldots, n\}$.

La familia de conjuntos $B \in \mathfrak{B}(\mathbb{R})$ tales que $I_1 \times \cdots \times I_{n-1} \times B \in \mathcal{H}$, para cualquier celda $I_1 \times \cdots \times I_{n-1} \in \mathbb{R}^{n-1}$ tal que $I_k \in \mathcal{G}$ para cualquier $k \in \{1, 2, \ldots, n-1\}$, forma una σ-álgebra que contiene a los intervalos de la familia $\{(-\infty, x] : x \in \mathbb{R}\}$; por lo tanto, contiene a todos los borelianos de $\mathbb{R}$.

De la misma manera, si B_n es un boreliano de $\mathbb{R}$ cualquiera, entonces la familia de conjuntos $B \in \mathfrak{B}(\mathbb{R})$ tales que $I_1 \times \cdots \times I_{n-2} \times B \times B_n \in \mathcal{H}$, para cualquier celda $I_1 \times \cdots \times I_{n-2}$ de $\mathbb{R}^{n-2}$ tal que $I_k \in \mathcal{G}$ para cualquier $k \in \{1, 2, \ldots, n-2\}$, forma una σ-álgebra que contiene a los intervalos de la familia $\{(-\infty, x] : x \in \mathbb{R}\}$; por lo tanto, contiene a todos los borelianos de $\mathbb{R}$.

Continuando con este procedimiento, se obtiene que los conjuntos de la forma $I \times B_2 \times \ldots \times B_n$, donde $I \in \mathcal{G}$ y $B_2, \ldots, B_n$ son borelianos cualesquiera de $\mathbb{R}$, pertenecen a $\mathcal{H}$.

Finalmente, si $B_2, \ldots, B_n$ son borelianos cualesquiera de $\mathbb{R}$, entonces la familia de conjuntos $B \in \mathfrak{B}(\mathbb{R})$ tales que $B \times B_2 \times \ldots \times B_n \in \mathcal{H}$, forma una σ-álgebra que contiene a los intervalos de la familia $\{(-\infty, x] : x \in \mathbb{R}\}$; por lo tanto, contiene a todos los borelianos de $\mathbb{R}$.

Así que, $\mathcal{H}$ contiene a todos los subconjuntos de $\mathbb{R}^n$ de la forma $B_1 \times B_2 \times \ldots \times B_n$, donde $B_1, \ldots, B_n$ son borelianos de $\mathbb{R}$.

En particular, $\mathcal{H}$ contiene a cualquier celda en $\mathbb{R}^n$, así que contiene a todos los borelianos de $\mathbb{R}^n$.

Finalmente, como $\mathcal{H} \subset \mathfrak{B}(\mathbb{R}^n)$, entonces $\mathcal{H} = \mathfrak{B}(\mathbb{R}^n)$. ∎

Corolario 2.2. *La σ-álgebra de Borel en $\mathbb{R}^n$ está generada por la familia de celdas de $\mathbb{R}^n$ de la forma $(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]$, donde $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.*

Corolario 2.3. *La σ-álgebra de Borel en $\mathbb{R}^n$ está generada por la familia de subconjuntos de $\mathbb{R}^n$ de la forma $B_1 \times B_2 \times \ldots \times B_n$, donde $B_1, \ldots, B_n$ son borelianos de $\mathbb{R}$.*

Proposición 2.5. *La σ-álgebra de Borel en $\mathbb{R}^n$ está generada por cualquiera de las siguientes familias de conjuntos:*

(i) $\mathcal{D}_1$: *Las celdas de $\mathbb{R}^n$ de la forma*
$(-\infty, x_1) \times (-\infty, x_2) \times \cdots \times (-\infty, x_n)$,
donde $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

(ii) $\mathcal{D}_2$: *Las celdas de $\mathbb{R}^n$ de la forma*

$(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n]$,
donde $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$.

(iii) $\mathcal{D}_3$*: Las celdas de* $\mathbb{R}^n$ *de la forma*
$(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)$,
donde $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$.

(iv) $\mathcal{D}_4$*: Las celdas de* $\mathbb{R}^n$ *de la forma*
$[a_1, b_1) \times [a_2, b_2) \times \cdots \times [a_n, b_n)$,
donde $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$.

(v) $\mathcal{D}_5$*: Las celdas de* $\mathbb{R}^n$ *de la forma*
$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$,
donde $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$.

Demostración

Denotemos por $\mathcal{D}$ a la familia de celdas de $\mathbb{R}^n$ de la forma:

$(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]$,

donde $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

a) Sea $(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n] \in \mathcal{D}$, entonces:

$(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]$

$= \bigcap_{m=1}^{\infty} \left(-\infty, x_1 + \frac{1}{m}\right) \times \left(-\infty, x_2 + \frac{1}{m}\right) \times \cdots \times \left(-\infty, x_n + \frac{1}{m}\right) \in \sigma(\mathcal{D}_1)$.

Así que, $\mathfrak{B}(\mathbb{R}^n) = \sigma(\mathcal{D}) \subset \sigma(\mathcal{D}_1)$.

Además, todo elemento de $\mathcal{D}_1$ es un boreliano de $\mathbb{R}^n$.

Por lo tanto, $\sigma(\mathcal{D}_1) = \mathfrak{B}(\mathbb{R}^n)$.

b) Sea $(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n] \in \mathcal{D}$, entonces:

$(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n] = \bigcup_{m=1}^{\infty} (-m, x_1] \times (-m, x_2] \times \cdots \times (-m, x_n] \in \sigma(\mathcal{D}_2)$.

Así que, $\mathfrak{B}(\mathbb{R}^n) = \sigma(\mathcal{D}) \subset \sigma(\mathcal{D}_2)$.

Además, todo elemento de $\mathcal{D}_2$ es un boreliano de $\mathbb{R}^n$.

Por lo tanto, $\sigma(\mathcal{D}_2) = \mathfrak{B}(\mathbb{R}^n)$.

c) Sea $(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n] \in \mathcal{D}_2$ una celda no vacía, entonces:

$(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n] = \bigcap_{m=1}^{\infty} \left(a_1, b_1 + \frac{1}{m}\right) \times \left(a_2, b_2 + \frac{1}{m}\right) \times \cdots \times \left(a_n, b_n + \frac{1}{m}\right) \in \sigma(\mathcal{D}_3)$.

Así que, $\mathfrak{B}(\mathbb{R}^n) = \sigma(\mathcal{D}_2) \subset \sigma(\mathcal{D}_3)$.

Además, todo elemento de $\mathcal{D}_3$ es un boreliano de $\mathbb{R}^n$.

Por lo tanto, $\sigma(\mathcal{D}_3) = \mathfrak{B}(\mathbb{R}^n)$.

d) Sea $(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n) \in \mathcal{D}_3$ una celda no vacía, entonces:

$(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n) = \bigcup_{m=1}^{\infty} \left[a_1 + \frac{1}{m}, b_1\right) \times \left[a_2 + \frac{1}{m}, b_2\right) \times \cdots \times \left[a_n + \frac{1}{m}, b_n\right) \in \sigma(\mathcal{D}_4)$.

Así que, $\mathfrak{B}(\mathbb{R}^n) = \sigma(\mathcal{D}_3) \subset \sigma(\mathcal{D}_4)$.

Además, todo elemento de $\mathcal{D}_4$ es un boreliano de $\mathbb{R}^n$.

Por lo tanto, $\sigma(\mathcal{D}_4) = \mathfrak{B}(\mathbb{R}^n)$.

e) Sea $[a_1, b_1) \times [a_2, b_2) \times \cdots \times [a_n, b_n) \in \mathcal{D}_4$ una celda no vacía, entonces:

$[a_1, b_1) \times [a_2, b_2) \times \cdots \times [a_n, b_n) = \bigcup_{m=1}^{\infty} \left[a_1, b_1 - \frac{1}{m}\right] \times \left[a_2, b_2 - \frac{1}{m}\right] \times \cdots \times \left[a_n, b_n - \frac{1}{m}\right] \in \sigma(\mathcal{D}_5)$.

Así que, $\mathfrak{B}(\mathbb{R}^n) = \sigma(\mathcal{D}_4) \subset \sigma(\mathcal{D}_5)$.

Además, todo elemento de $\mathcal{D}_5$ es un boreliano de $\mathbb{R}^n$.

Por lo tanto, $\sigma(\mathcal{D}_5) = \mathfrak{B}(\mathbb{R}^n)$.

■

Proposición 2.6. *La σ-álgebra de Borel en $\mathbb{R}^n$ está generada por la familia de subconjuntos abiertos de $\mathbb{R}^n$.*

Demostración

Sea G un subconjunto abierto no vacío de $\mathbb{R}^n$, entonces, para cada $x \in G$ existe una bola abierta B de centro x y radio $s > 0$ contenida en G.

Sea r un número racional positivo menor que s y $y = (y_1, y_2, \ldots, y_n)$ un elemento de la bola abierta de centro x y radio $\frac{1}{2n}r$ tal que, para cualquier $k \in \{1, 2, \ldots, n\}$, y_k es un número racional.

Obviamente x pertenece a la bola abierta de centro y y radio $\frac{1}{2n}r$, la cual está contenida en B.

Definamos:

$C = \left(y_1 - \frac{1}{2n}r, y_1 + \frac{1}{2n}r\right) \times \left(y_2 - \frac{1}{2n}r, y_2 + \frac{1}{2n}r\right) \times \cdots \times \left(y_n - \frac{1}{2n}r, y_n + \frac{1}{2n}r\right)$.

La distancia entre dos elementos cualesquiera de C es menor que la distancia entre los puntos $\left(y_1 - \frac{1}{2n}r, y_2 - \frac{1}{2n}r, \ldots, y_n - \frac{1}{2n}r\right)$ y $\left(y_1 + \frac{1}{2n}r, y_2 + \frac{1}{2n}r, \ldots, y_n + \frac{1}{2n}r\right)$, la cual es igual a $\frac{1}{\sqrt{n}}r$.

Como x pertenece a la bola abierta de centro y y radio $\frac{1}{2n}r$, si $x = (x_1, x_2, \ldots, x_n)$, entonces $|x_k - y_k| < \frac{1}{2n}r$ para cualquier $k \in \{1, 2, \ldots, n\}$, así que $x \in C$. Por lo tanto, si $z \in C$, entonces:

$d(z,x) < \frac{1}{\sqrt{n}} r \leq r < s$.

Así que $C \subset B \subset G$.

Denotemos por $\mathcal{C}$ al conjunto de celdas en $\mathbb{R}^n$ de la forma $(r_1, s_1) \times (r_2, s_2) \times \cdots \times (r_n, s_n)$, donde $r_1, s_1, r_2, s_2, \ldots, r_n, s_n$ son números racionales. $\mathcal{C}$ es entonces un conjunto numerable y, por lo anterior, para cada $x \in G$ existe $C \in \mathcal{C}$ tal que $x \in C$ y $C \subset G$.

Por lo tanto, G se puede expresar como la unión de una colección finita o infinita numerable de conjuntos en $\mathcal{C}$, cada uno de los cuales un boreliano de $\mathbb{R}^n$. Así que G es un boreliano de $\mathbb{R}^n$.

Finalmente, la familia de subconjuntos abiertos de $\mathbb{R}^n$ contiene a las celdas en $\mathbb{R}^n$ de la forma $(-\infty, x_1) \times (-\infty, x_2) \times \cdots \times (-\infty, x_n)$, donde $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, las cuales generan a la σ-álgebra de Borel en $\mathbb{R}^n$. Así que también la familia de subconjuntos abiertos de $\mathbb{R}^n$ genera a la σ-álgebra de Borel en $\mathbb{R}^n$.

■

Definición 2.11 (**σ-álgebra de Borel en $\overline{\mathbb{R}}^n$**). *La σ-álgebra de Borel en $\overline{\mathbb{R}}^n$, la cual será denotada por $\mathfrak{B}\left(\overline{\mathbb{R}}^n\right)$, es la σ-álgebra de subconjuntos de $\overline{\mathbb{R}}^n$ generada por $\overline{\mathcal{R}}$. A los elementos de esa σ-álgebra los llamaremos borelianos de $\overline{\mathbb{R}}^n$.*

Proposición 2.7. *Si $B_1, \ldots, B_n$ son borelianos de $\overline{\mathbb{R}}$, entonces $B_1 \times B_2 \times \ldots \times B_n$ es un boreliano de $\overline{\mathbb{R}}^n$.*

Demostración

Definamos $\mathcal{H} = \left\{(-\infty, x] : x \in \overline{\mathbb{R}}\right\} \cup \left\{\overline{\mathbb{R}}\right\}$

La familia de conjuntos $B \in \mathfrak{B}\left(\overline{\mathbb{R}}\right)$ tales que $I_1 \times \cdots \times I_{n-1} \times B \in \mathfrak{B}\left(\overline{\mathbb{R}}^n\right)$, para cualquier familia de intervalos $I_1 \times \cdots \times I_{n-1}$ tales que $I_k \in \mathcal{H}$ para cualquier $k \in \{1, 2, \ldots, n-1\}$, forma una σ-álgebra que contiene a los intervalos de la familia $\left\{(-\infty, x] : x \in \overline{\mathbb{R}}\right\}$; por lo tanto, contiene a todos los borelianos de $\overline{\mathbb{R}}$.

De la misma manera, si B_n es un boreliano de $\overline{\mathbb{R}}$ cualquiera, entonces la familia de conjuntos $B \in \mathfrak{B}\left(\overline{\mathbb{R}}\right)$ tales que $I_1 \times \cdots \times I_{n-2} \times B \times B_n \in \mathfrak{B}\left(\overline{\mathbb{R}}^n\right)$, para cualquier familia de intervalos $I_1 \times \cdots \times I_{n-2}$ tales que $I_k \in \mathcal{H}$ para cualquier $k \in \{1, 2, \ldots, n-2\}$, forma una σ-álgebra que contiene a los intervalos de la familia $\left\{(-\infty, x] : x \in \overline{\mathbb{R}}\right\}$; por lo tanto, contiene a todos los borelianos de $\overline{\mathbb{R}}$.

Continuando con este procedimiento, se obtiene que los conjuntos de la forma $I \times B_2 \times \ldots \times B_n$, donde $I \in \mathcal{H}$ y $B_2, \ldots, B_n$ son borelianos cualesquiera de $\overline{\mathbb{R}}$, son borelianos de $\overline{\mathbb{R}}^n$.

Finalmente, si $B_2, \ldots, B_n$ son borelianos cualesquiera de $\overline{\mathbb{R}}$, entonces la familia de conjuntos $B \in \mathfrak{B}\left(\overline{\mathbb{R}}\right)$ tales que $B \times B_2 \times \ldots \times B_n \in \mathfrak{B}\left(\overline{\mathbb{R}}^n\right)$, forma una σ-álgebra que contiene a los intervalos de la familia $\left\{(-\infty, x] : x \in \overline{\mathbb{R}}\right\}$; por lo tanto, contiene a todos los borelianos de $\overline{\mathbb{R}}$.

■

Corolario 2.4. *La σ-álgebra de Borel en $\overline{\mathbb{R}}^n$ está generada por la familia de conjuntos de la forma $B_1 \times B_2 \times \ldots \times B_n$, donde $B_1, \ldots, B_n$ son borelianos de $\overline{\mathbb{R}}$.*

Proposición 2.8. *La σ-álgebra de Borel en $\overline{\mathbb{R}}^n$ está generada por la familia de subconjuntos de $\overline{\mathbb{R}}^n$ de la forma $I_1 \times I_2 \times \cdots \times I_n$, donde, para cualquier $k \in \{1, 2, \ldots, n\}$, I_k es ya sea un intervalo de la forma $[-\infty, x]$, con $x \in \mathbb{R}$, o bien $I_k = \overline{\mathbb{R}}$.*

Demostración

Denotemos por $\overline{\mathcal{R}}'$ a la familia de subconjuntos de $\overline{\mathbb{R}}^n$ de la forma $I_1 \times I_2 \times \cdots \times I_n$, donde, para cualquier $k \in \{1, 2, \ldots, n\}$, I_k es ya sea un intervalo de la forma $[-\infty, x]$, con $x \in \mathbb{R}$, o bien $I_k = \overline{\mathbb{R}}$.

Sean $R = I_1 \times I_2 \times \cdots \times I_n \in \overline{\mathcal{R}}$ una celda no vacía y $U \subset \{1, 2, \ldots, n\}$ tal que $I_k \neq \overline{\mathbb{R}}$ si $k \in U$ e $I_k = \overline{\mathbb{R}}$ si $k \notin U$. Para cada $j \in \{1, 2, \ldots, n\}$, $k \in U$ y $m \in \mathbb{N}$, definamos:

$$A_j = \begin{cases} [-\infty, x_j] & \text{si } j \in U \text{ e } I_j = (-\infty, x_j] \\ \overline{\mathbb{R}} & \text{si } j \notin U \end{cases}$$

$$B_{jk} = \begin{cases} [-\infty, x_j] & \text{si } j \in U,\ j \neq k \text{ e } I_j = (-\infty, x_j] \\ \{-\infty\} & \text{si } j = k \\ \overline{\mathbb{R}} & \text{si } j \notin U \end{cases}$$

$$B_{jk}^{(m)} = \begin{cases} [-\infty, x_j] & \text{si } j \in U,\ j \neq k \text{ e } I_j = (-\infty, x_j] \\ [-\infty, -m] & \text{si } j = k \\ \overline{\mathbb{R}} & \text{si } j \notin U \end{cases}$$

$$R_k = B_{1k} \times B_{2k} \times \cdots \times B_{nk}$$

Entonces:

$$R_k = \bigcap_{m=1}^{\infty} B_{1k}^{(m)} \times B_{2k}^{(m)} \times \cdots \times B_{nk}^{(m)} \in \sigma\left(\overline{\mathcal{R}}'\right).$$

Así que:

$$R = A_1 \times A_2 \times \cdots \times A_n - \bigcup_{k \in U} R_k \in \sigma\left(\overline{\mathcal{R}}'\right).$$

■

Corolario 2.5. *La σ-álgebra de Borel en $\overline{\mathbb{R}}^n$ está generada por la familia de subconjuntos de $\overline{\mathbb{R}}^n$ de la forma $[-\infty, x_1] \times \cdots \times [-\infty, x_n]$, donde $x_k \in \overline{\mathbb{R}}$ para cualquier $k \in \{1, 2, \ldots, n\}$.*

Proposición 2.9. *La σ-álgebra de Borel en $\overline{\mathbb{R}}^n$ está generada por la familia de subconjuntos de $\overline{\mathbb{R}}^n$ de la forma $I_1 \times I_2 \times \cdots \times I_n$, donde, para cualquier $k \in \{1, 2, \ldots, n\}$, I_k es ya sea un intervalo de la forma $[-\infty, x)$, con $x \in \mathbb{R}$, o bien $I_k = \overline{\mathbb{R}}$.*

Demostración

Denotemos por $\overline{\mathcal{R}}'$ a la familia de subconjuntos de $\overline{\mathbb{R}}^n$ de la forma $I_1 \times I_2 \times \cdots \times I_n$, donde, para cualquier $k \in \{1, 2, \ldots, n\}$, I_k es ya sea un intervalo de la forma $[-\infty, x]$, con $x \in \mathbb{R}$, o bien $I_k = \overline{\mathbb{R}}$, y por $\overline{\mathcal{R}}''$ a la familia de subconjuntos de $\overline{\mathbb{R}}^n$ de la forma $I_1 \times I_2 \times \cdots \times I_n$, donde, para cualquier $k \in \{1, 2, \ldots, n\}$, I_k es ya sea un intervalo de la forma $[-\infty, x)$, con $x \in \mathbb{R}$, o bien $I_k = \overline{\mathbb{R}}$.

Sean $R = I_1 \times I_2 \times \cdots \times I_n \in \overline{\mathcal{R}}'$ y $U \subset \{1, 2, \ldots, n\}$ tal que $I_j \neq \overline{\mathbb{R}}$ si $j \in U$ e $I_j = \overline{\mathbb{R}}$ si $j \notin U$. Para cada $k \in \{1, 2, \ldots, n\}$ y $m \in \mathbb{N}$, definamos:

$$J_k^m = \begin{cases} \left[-\infty, x_k + \frac{1}{m}\right) & \text{si } I_k = [-\infty, x_k] \text{ y } k \in U \\ \overline{\mathbb{R}} & \text{si } j \notin U \end{cases}$$

Entonces:

$$R = \bigcap_{m=1}^{\infty} J_1^{(m)} \times J_2^{(m)} \times \cdots \times J_n^{(m)} \in \sigma\left(\overline{\mathcal{R}}''\right).$$

Sean $R = I_1 \times I_2 \times \cdots \times I_n \in \overline{\mathcal{R}}''$ y $U \subset \{1, 2, \ldots, n\}$ tal que $I_j \neq \overline{\mathbb{R}}$ si $j \in U$ e $I_j = \overline{\mathbb{R}}$ si $j \notin U$. Para cada $k \in \{1, 2, \ldots, n\}$ y $m \in \mathbb{N}$, definamos:

$$J_k^m = \begin{cases} \left[-\infty, x_k - \frac{1}{m}\right] & \text{si } I_k = [-\infty, x_k) \text{ y } k \in U \\ \overline{\mathbb{R}} & \text{si } j \notin U \end{cases}$$

Entonces:

$$R = \bigcup_{m=1}^{\infty} J_1^{(m)} \times J_2^{(m)} \times \cdots \times J_n^{(m)} \in \sigma\left(\overline{\mathcal{R}}'\right).$$

■

2.3. Funciones finitamente aditivas y σ-aditivas

Corolario 2.6. *La σ-álgebra de Borel en $\overline{\mathbb{R}}$ está generada por la familia de intervalos de la forma $[-\infty, x)$, donde $x \in \overline{\mathbb{R}}$.*

Definición 2.12 (**Función finitamente aditiva sobre un álgebra**). *Sea $\mathbb{F}$ un conjunto y $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$. Se dice que una función no negativa $\mu : \mathcal{A} \mapsto \overline{\mathbb{R}}$ es finitamente aditiva si dada cualquier familia finita, $A_1, \ldots, A_n$, de elementos de $\mathcal{A}$ tal que $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces $\mu(\bigcup\limits_{k=1}^{n} A_k) = \sum_{k=1}^{n} \mu(A_k)$.*

Obsérvese que, si $\mathcal{A}$ es un álgebra de subconjuntos de $\mathbb{F}$, para probar que una función $\mu : \mathcal{A} \mapsto \overline{\mathbb{R}}$ es finitamente aditiva, basta con demostrar que si A y B son dos conjuntos ajenos del álgebra, entonces $\mu(A \cup B) = \mu(A) + \mu(B)$. Teniendo esta propiedad, la aditividad finita se prueba con un razonamiento de inducción.

Definición 2.13 (**Función $\boldsymbol{\sigma}$-aditiva sobre un álgebra**). *Sea $\mathbb{F}$ un conjunto y $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$. Se dice que una función no negativa $\mu : \mathcal{A} \mapsto \overline{\mathbb{R}}$ es σ-aditiva si es finitamente aditiva y dada cualquier familia infinita numerable, $A_1, A_2, \ldots$, de elementos de $\mathcal{A}$ tal que $A_i \cap A_j = \emptyset$ para $i \neq j$ y $\bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{A}$, entonces $\mu\left(\bigcup\limits_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$.*

Definición 2.14 (**Función $\boldsymbol{\sigma}$-aditiva sobre una σ-álgebra**). *Sea $\mathbb{F}$ un conjunto y $\Im$ una σ-álgebra de subconjuntos de $\mathbb{F}$. Se dice que una función no negativa $\mu : \Im \mapsto \overline{\mathbb{R}}$ es σ-aditiva si es finitamente aditiva y dada cualquier familia infinita numerable, $A_1, A_2, \ldots$, de elementos de $\Im$ tal que $A_i \cap A_j = \emptyset$, entonces $\mu\left(\bigcup\limits_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$.*

Proposición 2.10. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ una función no negativa finitamente aditiva, entonces:*

(i) *Si $A, B \in \mathcal{A}$ y $A \subset B$, entonces $\mu(A) \leq \mu(B)$.*
(ii) *Si $A_1, \ldots, A_n \in \mathcal{A}$, entonces $\mu\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} \mu(A_k)$.*

Demostración

Sean $A, B \in \mathcal{A}$ tales que $A \subset B$, entonces $B = A \cup (B - A)$, así que $\mu(B) = \mu(A) + \mu(B - A)$. Por lo tanto, $\mu(A) \leq \mu(B)$, ya que μ es no negativa.

Sean ahora $A_1, \ldots, A_n \in \mathcal{A}$ y definamos $A_0 = \emptyset$, entonces:

$$\mu\left(\bigcup_{k=1}^{n} A_k\right) = \mu\left(\bigcup_{k=1}^{n}\left(A_k - \bigcup_{j=0}^{k-1} A_j\right)\right)$$

$$= \sum_{k=1}^{n} \mu\left(A_k - \bigcup_{j=0}^{k-1} A_j\right) \leq \sum_{k=1}^{n} \mu(A_k).$$

■

Corolario 2.7. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ una función no negativa finitamente aditiva, entonces:*

Si $A, B \in \mathcal{A}$, $A \subset B$ y $\mu(A) < \infty$, entonces:

$\mu(B - A) = \mu(B) - \mu(A)$.

Para cualquier pareja $A, B \in \mathcal{A}$ tal que $\mu(A) < \infty$ o $\mu(B) < \infty$, se tiene:

$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

Proposición 2.11. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y $\mu : \mathcal{A} \to \mathbb{R}$ una función no negativa finitamente aditiva, entonces, para cualquier colección finita $A_1, A_2, \ldots, A_n$ de elementos de $\mathcal{A}$, se tiene:*

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k) - \sum_{\{i,j \in \{1,\ldots,n\},:i\neq j\}} P(A_i \cap A_j)$$

$$+ \sum_{\{i,j,k \in \{1,\ldots,n\},:i\neq j, j\neq k, i\neq k\}} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P(A_1 \cap A_2 \ldots \cap A_n).$$

Demostración

La demostración se hará por inducción sobre el número de eventos n. Para $n = 2$ ya se tiene el resultado. Supongamos ahora que la propiedad es válida para el caso de cualesquiera $m-1$ eventos y sean $A_1, \ldots A_m$ m eventos cualesquiera. Se tiene entonces:

$$P(\cup_{k=1}^{m} A_k) = P(\cup_{k=1}^{m-1} A_k) + P(A_m) - P(A_m \cap \{\cup_{k=1}^{m-1} A_k\})$$

$$= P(\cup_{k=1}^{m-1} A_k) + P(A_m) - P(\cup_{k=1}^{m-1} A_k \cap A_m)$$

$$= \sum_{k=1}^{m-1} P(A_k) - \sum_{\{i,j\in\{1,\ldots m-1\}, i\neq j\}} P(A_i \cap A_j)$$

$$+ \sum_{\{i,j,k\in\{1,\ldots m-1\}, i\neq j, j\neq k, i\neq k\}} P(A_i \cap A_j \cap A_k) - \ldots$$

$$+(-1)^m P(A_1 \cap A_2 \ldots \cap A_{m-1}) + P(A_m) - \sum_{k=1}^{m-1} P(A_k \cap A_n)$$

$$+ \sum_{\{i,j\in\{1,\ldots m-1\}, i\neq j\}} P((A_i \cap A_m) \cap (A_j \cap A_m))$$

$$- \sum_{\{i,j,k\in\{1,\ldots m-1\}, i\neq j, j\neq k, i\neq k\}} P((A_i \cap A_m) \cap (A_j \cap A_m) \cap (A_k \cap A_m))$$

$$+ \ldots - (-1)^m P((A_1 \cap A_m) \cap (A_2 \cap A_m) \ldots \cap (A_{m-1} \cap A_m))$$

$$= \sum_{k=1}^{m} P(A_k) - \sum_{\{i,j\in\{1,\ldots m\}, i\neq j\}} P(A_i \cap A_j)$$

$$+ \sum_{\{i,j,k\in\{1,\ldots m\}, i\neq j, j\neq k, i\neq k\}} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{m+1} P(A_1 \cap \ldots \cap A_m).$$

Por lo tanto, la propiedad es válida para m eventos cualesquiera.

Así que, por el principio de inducción matemática, la propiedad es válida para cualquier $n \in \mathbb{N}$.

■

TEOREMA 2.1. *Sea $\mathbb{F}$ un conjunto, $\Im$ una σ-álgebra de subconjuntos de $\mathbb{F}$ y $\mu : \Im \to \overline{\mathbb{R}}$ función no negativa σ-aditiva. Entonces:*

(i) *Para cualquier sucesión creciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\Im$, se tiene:*
$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$.
(ii) *Para cualquier sucesión decreciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\Im$, tales que $\mu(A_N) < \infty$ para alguna $N \in \mathbb{N}$, se tiene:*
$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\rightsquigarrow\infty} \mu(A_n)$.

Demostración

1. Sea $(A_n)_{n\in\mathbb{N}}$ sucesión creciente de elementos de $\Im$.

Si $\mu(A_n) = \infty$ para alguna $n \in \mathbb{N}$, entonces $\lim_{n\rightsquigarrow\infty} \mu(A_n) = \infty$ y $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \infty$; así que $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\rightsquigarrow\infty} \mu(A_n)$.

Supongamos ahora que $\mu(A_n) < \infty$ para cualquier $n \in \mathbb{N}$. Definamos $B_1 = A_1$ y, para cada $n \in \{2, 3, \ldots\}$, $B_n = A_n - A_{n-1}$. Entonces los conjuntos $B_1, B_2, \ldots$ pertenecen a $\Im$, son ajenos por parejas y $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Así que:

$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) = \text{lím}_{n\rightsquigarrow\infty} \sum_{k=1}^{n} \mu(B_k)$

$= \mu(B_1) + \text{lím}_{n\rightsquigarrow\infty} \sum_{k=2}^{n} \mu(A_k - A_{k-1}) = \mu(B_1) + \text{lím}_{n\rightsquigarrow\infty} \sum_{k=2}^{n} [\mu(A_k) - \mu(A_{k-1})]$

$= \mu(B_1) + \text{lím}_{n\rightsquigarrow\infty} \mu(A_n) - \mu(A_1) = \text{lím}_{n\rightsquigarrow\infty} \mu(A_n)$.

2. Sea $(A_n)_{n\in\mathbb{N}}$ sucesión decreciente de elementos de $\Im$ tales que $\mu(A_N) < \infty$ para alguna $N \in \mathbb{N}$.

Para cada $k \in \{N+1, N+2, \ldots\}$, definamos $B_k = A_N - A_k$. Entonces la sucesión $(B_{N+n})_{n\in\mathbb{N}}$ es creciente y $\bigcup_{n=N+1}^{\infty} B_n = A_N - \bigcap_{n=N+1}^{\infty} A_n$, así que:

$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \mu\left(\bigcap_{n=N+1}^{\infty} A_n\right) = \mu(A_N) - \mu\left(\bigcup_{n=N+1}^{\infty} B_n\right)$

$= \mu(A_N) - \text{lím}_{n\rightsquigarrow\infty} \mu(B_{N+n}) = \mu(A_N) - \text{lím}_{n\rightsquigarrow\infty} \mu(A_N - A_{N+n})$

$= \text{lím}_{n\to\infty} \mu(A_{N+n}) = \text{lím}_{n\rightsquigarrow\infty} \mu(A_n)$.

$\blacksquare$

2.4. La medida de Lebesgue en $\mathbb{R}$

En esta sección expondremos la formulación moderna de la teoría de la medida de Lebesgue en $\mathbb{R}$, de manera que todos los conjuntos con los que trataremos serán subconjuntos de los números reales.

La medida de Lebesgue se define mediante un proceso de extensión partiendo de que la medida de un intervalo es su longitud. Se define la medida exterior de cualquier subconjunto de $\mathbb{R}$ acercándonos a ese conjunto mediante uniones numerables de intervalos abiertos que contienen al conjunto. No se define una medida interior, como lo hizo Lebesgue, ya que la caracterización que hizo Carathéodory de los conjuntos medibles es más cómoda de trabajar y requiere únicamente contar con una medida exterior

Definición 2.15. *Diremos que una colección finita o infinita numerable de intervalos abiertos finitos $I_1, I_2, \ldots$ es una cubierta del conjunto A si $A \subset \bigcup_n I_n$.*

El resultado siguiente, cuya demostración es muy simple gracias al teorema de Heine-Borel, es fundamental para demostrar que la medida de Lebesgue de un intervalo es igual a su longitud. De esta forma, lo que obtenemos es una extensión a todos los conjuntos medibles del concepto de longitud.

Lema 2.1. *Sea I un intervalo finito de cualquier tipo (i.e. abierto, semiabierto, etc.) e $I_1, I_2, \ldots$ una cubierta de I, entonces:*

$l(I) \leq \sum_j l(I_j)$.

Demostración

Sean a y b los extremos del intervalo I y, dada $\varepsilon > 0$, sean I_a e I_b intervalos abiertos que contengan a a y b respectivamente y tales que $l(I_a) = l(I_b) = \frac{\varepsilon}{2}$. Entonces los intervalos $I_a, I_b, I_1, I_2, \ldots$ forman una cubierta del intervalo $[a, b]$. Por el teorema de Heine-Borel, existe entonces una subcubierta finita. Sea L la suma de las longitudes de los intervalos de dicha cubierta finita, entonces:

$l(I) = b - a \leq L \leq \sum_j l(I_j) + \varepsilon$.

Se tiene entonces que $l(I) \leq \sum_j l(I_j) + \varepsilon$ para cualquier $\varepsilon > 0$, de lo cual se sigue el resultado.

■

Definición 2.16. *Se define la medida exterior, $m_e(A)$, de un conjunto A, mediante la relación:*

$m_e(A) = \text{ínf}\left\{\sum_j l(I_j) : I_1, I_2, \ldots \text{ es cubierta de } A\right\}$.

Proposición 2.12. *Si A y B son dos conjuntos tales que $A \subset B$ entonces:*

$m_e(A) \leq m_e(B)$.

Proposición 2.13. *La medida exterior de un intervalo es igual a su longitud.*

Demostración

Consideremos primero un intervalo finito I. Por el lema 2.1 se tiene $l(I) \leq m_e(I)$.

Dada $\varepsilon > 0$ sea J un intervalo abierto tal que $J \supset I$ y $l(J) < l(I) + \varepsilon$. Como $J \supset I$, J es cubierta de I, así que se tiene:

$m_e(I) \leq l(J) < l(I) + \varepsilon$.

Es decir, $m_e(I) < l(I) + \varepsilon$ para cualquier $\varepsilon > 0$. Por lo tanto, $m_e(I) \leq l(I)$.

Si el intervalo I es infinito, dado cualquier $\alpha > 0$ existe un intervalo finito J contenido en I y de longitud α. Por lo tanto, $m_e(I) \geq m_e(J) = l(J) = \alpha$. Así que $m_e(I) = \infty$.

■

Ahora viene la propiedad que caracteriza a una medida exterior.

Proposición 2.14. *Si $A_1, A_2, \ldots$ es una colección finita o infinita numerable de conjuntos, entonces:*

$m_e\left(\bigcup_n A_n\right) \leq \sum_n m_e(A_n)$.

Demostración

Si $m_e(A_n) = \infty$ para alguna n el resultado es trivial.

Supongamos entonces que $m_e(A_n) < \infty$ para toda n. Dada $\varepsilon > 0$, para cada conjunto A_n sea $I_{n,1}, I_{n,2}, \ldots$ una cubierta de A_n tal que $\sum_m l(I_{n,m}) < m_e(A_n) + \frac{\varepsilon}{2^n}$. La familia de intervalos $I_{n,m}$ forman una cubierta de $\bigcup_n A_n$, así que:

$m_e\left(\bigcup_n A_n\right) \leq \sum_n \sum_m l(I_{n,m}) \leq \sum_n \left[m_e(A_n) + \frac{\varepsilon}{2^n}\right] \leq \sum_n m_e(A_n) + \varepsilon.$

Es decir, $m_e\left(\bigcup_n A_n\right) \leq \sum_n m_e(A_n) + \varepsilon$ para cualquier $\varepsilon > 0$. Por lo tanto:

$m_e\left(\bigcup_n A_n\right) \leq \sum_n m_e(A_n).$

∎

Definición 2.17. *La propiedad enunciada en la última proposición es llamada la propiedad de σ-subaditividad de la medida exterior.*

Como decíamos, seguimos ahora el método de Carathéodory para definir la medibilidad de un conjunto.

Definición 2.18. *Se dice que un conjunto E es Lebesgue medible si:*

$m_e(A) = m_e(A \cap E) + m_e(A \cap E^c)$

para cualquier conjunto A. Además, en ese caso, se define la medida de E, $m(E)$, como la medida exterior de E.

Obsérvese que, por la σ-subaditividad de la medida exterior, se tiene:

$$m_e(A) \leq m_e(A \cap E) + m_e(A \cap E^c)$$

para cualquier par de conjuntos E y A, de manera que para demostrar la medibilidad de un conjunto E únicamente es necesario probar la otra desigualdad.

Ahora se trata de demostrar que la familia de conjuntos medibles forma una σ-álgebra de subconjuntos de $\mathbb{R}$ y que la función que asigna a cada conjunto medible su medida es σ-aditiva. Comenzamos demostrando primero que forma un álgebra y que la función medida es finitamente aditiva.

Proposición 2.15. *La familia de conjuntos Lebesgue medibles forma un álgebra de subconjuntos de $\mathbb{R}$.*

Demostración

Que $\mathbb{R}$ es medible, así como que el complemento de un conjunto medible es medible, son resultados obvios.

Sean E_1 y E_2 dos conjuntos medibles y A cualquier conjunto. Se tiene entonces:

$m_e\left(A \cap (E_1 \cup E_2)\right) + m_e\left(A \cap (E_1 \cup E_2)^c\right)$

$= m_e\left((A \cap E_1) \cup (A \cap E_1^c \cap E_2)\right) + m_e\left(A \cap E_1^c \cap E_2^c\right)$

$\leq m_e(A \cap E_1) + m_e(A \cap E_1^c \cap E_2) + m_e\left(A \cap E_1^c \cap E_2^c\right)$

$= m_e(A \cap E_1) + m_e(A \cap E_1^c) = m_e(A)$.

Así que, $E_1 \cup E_2$ es medible.

■

PROPOSICIÓN 2.16. *Sea $E_1, E_2, \ldots, E_n$ cualquier colección finita de conjuntos Lebesgue medibles, ajenos por parejas, entonces:*

$m_e\left(A \cap (\bigcup_{j=1}^{n} E_j)\right) = \sum_{j=1}^{n} m_e(A \cap E_j)$

para cualquier conjunto A.

Demostración

Para $n = 1$ la igualdad es obvia.

Supongamos ahora que la igualdad es válida para $n = k$ y sea $E_1, E_2, \ldots, E_{k+1}$ una colección finita de $k+1$ conjuntos medibles, ajenos por parejas, entonces, como E_{k+1} es Lebesgue medible, se tiene:

$m_e\left(A \cap (\bigcup_{j=1}^{k+1} E_j)\right) = m_e\left(A \cap (\bigcup_{j=1}^{k+1} E_j) \cap E_{k+1}\right) + m_e\left(A \cap (\bigcup_{j=1}^{k+1} E_j) \cap E_{k+1}^c\right)$

$= m_e\left(A \cap E_{k+1}\right) + m_e\left(A \cap (\bigcup_{j=1}^{k} E_j)\right)$

$= m_e\left(A \cap E_{k+1}\right) + \sum_{j=1}^{k} m_e(A \cap E_j) = \sum_{j=1}^{k+1} m_e(A \cap E_j)$.

■

COROLARIO 2.8. *La función que asigna a cada conjunto Lebesgue medible E su medida, $m(E)$, es una función finitamente aditiva.*

PROPOSICIÓN 2.17. *La familia de conjuntos Lebesgue medibles forma una σ-álgebra de subconjuntos de $\mathbb{R}$.*

Demostración

Sea $E_1, E_2, \ldots$ una colección infinita numerable de conjuntos Lebesgue medibles, ajenos por parejas. Como la familia de conjuntos Lebesgue medibles forma un álgebra de subconjuntos de $\mathbb{R}$, para cada $n \in \mathbb{N}$ el conjunto $\bigcup_{j=1}^{n} E_j$ es Lebesgue medible, así que, utilizando la proposición 5.5 y la σ-subaditividad de la medida exterior, se tiene, para cualquier conjunto A:

$m_e(A) = m_e\left(A \cap (\bigcup_{j=1}^{n} E_j)\right) + m_e\left(A \cap (\bigcup_{j=1}^{n} E_j)^c\right)$

$= \sum_{j=1}^{n} m_e(A \cap E_j) + m_e\left(A \cap (\bigcup_{j=1}^{n} E_j)^c\right)$

$\geq \sum_{j=1}^{n} m_e(A \cap E_j) + m_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)^c\right)$.

Tomando límite cuando $n \rightsquigarrow \infty$, se obtiene:

$m_e(A) \geq \sum_{j=1}^{\infty} m_e(A \cap E_j) + m_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)^c\right)$

$\geq m_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)\right) + m_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)^c\right)$.

Por lo tanto, $\bigcup_{j=1}^{\infty} E_j$ es Lebesgue medible.

■

Proposición 2.18. *La función que asigna a cada conjunto Lebesgue medible E su medida, $m(E)$, es una función σ-aditiva.*

Demostración

Sea $E_1, E_2, \ldots$ una colección infinita numerable de conjuntos Lebesgue medibles, ajenos por parejas. Por la σ-subaditividad de la medida exterior, se tiene $m(\bigcup_{j=1}^{\infty} E_j) \leq \sum_{j=1}^{\infty} m(E_j)$. Por otra parte, por la aditividad finita de la función que asigna a cada conjunto Lebesgue medible su medida, se tiene, para cualquier $n \in \mathbb{N}$:

$m(\bigcup_{j=1}^{\infty} E_j) \geq m(\bigcup_{j=1}^{n} E_j) = \sum_{j=1}^{n} m(E_j)$,

así que tomando límite cuando $n \rightsquigarrow \infty$, se tiene:

$m(\bigcup_{j=1}^{\infty} E_j) \geq \sum_{j=1}^{\infty} m(E_j)$

■

El resultado siguiente, cuya demostración es muy simple, es importante ya que hace ver que la familia de conjuntos medibles no está formada únicamente por los conjuntos borelianos de $\mathbb{R}$. Es la parte que la faltó considerar a Borel cuando desarrollo su teoría de la medida, ya que quedándonos únicamente con la medida de los borelianos, la medida no resulta ser una generalización del contenido de Jordan. La familia de conjuntos Jordan-medibles (aquellos cuyo contenido interior coincide con su contenido exterior) contiene a todos los conjuntos de contenido cero; por ejemplo, el conjunto de Cantor es Jordan-medible ya que su contenido es cero; por lo tanto, todo subconjunto del conjunto de Cantor es Jordan-medible. Como el conjunto de Cantor tiene la misma cardinalidad que el conjunto de números reales, resulta enconces que la cardinalidad de la familia de conjuntos que son Jordan-medibles coincide con la cardinalidad de la familia formada por todos los subconjuntos de $\mathbb{R}$. La familia de los conjuntos borelianos, en cambio, tiene la misma cardinalidad que $\mathbb{R}$.

EJERCICIO 2.4. *Demuestra que todo conjunto de medida exterior cero es Lebesgue medible.*
PROPOSICIÓN 2.19. *Todo intervalo es Lebesgue medible.*

Demostración

Como la familia de conjuntos Lebesgue medibles forma una σ-álgebra de subconjuntos de $\mathbb{R}$ y los conjuntos formados por un punto son Lebesgue medibles, es suficiente con probar que los intervalos de la forma $[a, \infty)$ son Lebesgue medibles.

Sea E un intervalo de la forma $[a, \infty)$, A cualquier conjunto e $I_1, I_2, \ldots$ una cubierta de A, entonces, para cada I_n, los conjuntos $I_n \cap E$ e $I_n \cap E^c$ son intervalos y se tiene:

$m_e(A \cap E) \leq m_e\left(\left(\bigcup_n I_n\right) \cap E\right) = m_e\left(\bigcup_n (I_n \cap E)\right)$

$\leq \sum_n m_e(I_n \cap E) = \sum_n l(I_n \cap E)$

$m_e(A \cap E^c) \leq m_e\left(\left(\bigcup_n I_n\right) \cap E^c\right) = m_e\left(\bigcup_n (I_n \cap E^c)\right)$

$\leq \sum_n m_e(I_n \cap E^c) = \sum_n l(I_n \cap E^c)$.

Así que:

$m_e(A \cap E) + m_e(A \cap E^c) \leq \sum_n l(I_n \cap E) + \sum_n l(I_n \cap E^c) = \sum_n l(I_n)$.

Finalmente, como lo anterior es válido para cualquier cubierta de A, se puede concluir que:

$m_e(A \cap E) + m_e(A \cap E^c) \leq m_e(A)$.

■

EJERCICIO 2.5. *Demuestra que todo conjunto boreliano es Lebesgue medible.*

El resultado siguiente muestra que la familia de los conjuntos medibles, si bien es más grande que la familia de los conjuntos borelianos, está formada por los conjuntos que difieren de un boreliano únicamente por un conjunto de medida cero. Es decir, no hay algún conjunto medible que no se obtenga de un boreliano ya sea quitándole o agregándole un conjunto de medida cero.

PROPOSICIÓN 2.20. *Dado cualquier conjunto Lebesgue medible E existe un boreliano B y un conjunto C de medida cero tales que $E = B \cup C$ y $B \cap C = \emptyset$.*

Demostración

Consideremos primero el caso de un conjunto Lebesgue medible E contenido en un intervalo finito (a, b) y definamos $F = (a, b) - E$.

La idea es cubrir F con un boreliano A tal que $m(A - F) = 0$, después de lo cual nos acercamos a E por dentro mediante $(a, b) - A$.

Dada $\varepsilon > 0$, sea $I_1, I_2, \ldots$ una cubierta de F tal que:

$\sum_j l(I_j) < m(F) + \varepsilon$.

$A^{(\varepsilon)} = \bigcup_j I_j$ es entonces un boreliano tal que:

$m(A^{(\varepsilon)} - F) = m(A^{(\varepsilon)}) - m(F) \leq \sum_j l(I_j) - m(F) < \varepsilon$.

Es decir, dada $\varepsilon > 0$ existe un boreliano $A^{(\varepsilon)}$ tal que $A^{(\varepsilon)} \supset F$ y $m(A^{(\varepsilon)} - F) < \varepsilon$.

Sea entonces $\left(A^{(n)}\right)_{n\in\mathbb{N}}$ una sucesión decreciente de borelianos que contengan a F y tales que:

$m(A^{(n)} - F) < \frac{1}{n}$

para cualquier $n \in \mathbb{N}$.

Se tiene entonces $m\left(\bigcap_{j=1}^{\infty} A^{(j)} - F\right) \leq m(A^{(n)} - F) < \frac{1}{n}$ para cualquier $n \in \mathbb{N}$, así que:

$m\left(\bigcap_{j=1}^{\infty} A^{(j)} - F\right) = 0$.

Por lo tanto, $A = (a,b) \cap \left(\bigcap_{j=1}^{\infty} A^{(j)}\right)$ es un boreliano tal que $A \supset F$ y $m(A - F) = 0$.

Definamos $B = (a,b) - A$, entonces B es boreliano y se tiene:

$B = (E \cup F) \cap A^c = (E \cap A^c) \cup (F \cap A^c) = E \cap A^c$,

$E \cap A = ((a,b) - F) \cap A = A - F$.

Así que, definiendo $C = A - F$, se tiene $E = B \cup C$, $B \cap C = \emptyset$ y $m(C) = 0$.

Tomemos ahora un conjunto Lebesgue medible E arbitrario y, para cada $k \in \mathbb{N}$, definamos $E_k = E \cap (-k,k)$.

Sea B_k un boreliano y C_k un conjunto de medida cero tales que $E_k = B_k \cup C_k$ y $B_k \cap C_k = \emptyset$, entonces tomando $B = \bigcup_{k=1}^{\infty} B_k$ y $D = \bigcup_{k=1}^{\infty} C_k$, se tiene que B es boreliano, D tiene medida cero y $E = B \cup D$.

Finalmente, definamos $C = D - B$, entonces C tiene medida cero y se tiene $E = B \cup C$ y $B \cap C = \emptyset$.

∎

Como corolario, se tiene el siguiente resultado:

TEOREMA 2.2. *La σ-álgebra de los conjuntos Lebesgue medibles es la σ-álgebra generada por los intervalos y los conjuntos de medida exterior cero.*

Obsérvese que la familia de los conjuntos de medida exterior cero coindice con la familia de los conjuntos de medida cero de acuerdo con la definición de Borel: **Un conjunto tiene**

medida cero si para cualquier $\varepsilon > 0$ existe un conjunto finito o infinito numerable de intervalos abiertos, cuya unión contiene al conjunto dado, y tales que la suma de sus longitudes es menor que ε. Así que los resultados anteriores pueden condensarse en el siguiente:

TEOREMA 2.3. *Existe una medida definida sobre la σ-álgebra de subconjuntos de $\mathbb{R}$ generada por los intervalos y los conjuntos de medida cero tal que la medida de cualquier intervalo es igual a su longitud.*

El siguiente resultado muestra lo cerca que estaba Borel de la teoría de la medida de Lebesgue. Bastaba con agregar a los borelianos los subconjuntos de los borelianos de medida cero y generar una σ-álgebra con esos conjuntos.

PROPOSICIÓN 2.21. *Todo conjunto de medida exterior cero está contenido en un conjunto boreliano de medida exterior cero.*

Demostración

Sea A un conjunto de medida exterior cero. Para cada $n \in \mathbb{N}$, sea $I_1^n, I_2^n \ldots$ una colección finita o infinita numerable de intervalos abiertos tales que $A \subset \bigcup_{k=1} I_k^n$ y $\sum_{k=1} l(I_k^n) < \frac{1}{n}$. Definamos $B_n = \bigcup_{k=1} I_k^n$ y $B = \bigcap_{n=1}^{\infty} B_n$. Entonces, $B \in \mathfrak{B}(\mathbb{R})$, tiene medida exterior cero y $A \subset B$.

■

DEFINICIÓN 2.19. *Denotaremos por $\mathfrak{L}(\mathbb{R})$ a la σ-álgebra formada por los conjuntos Lebesgue medibles en $\mathbb{R}$ y por λ a la medida de Lebesgue sobre $\mathfrak{L}(\mathbb{R})$.*

Obsérvese que la σ-álgebra de subconjuntos de $\overline{\mathbb{R}}$ generada por $\mathfrak{L}(\mathbb{R})$ y los conjuntos $\{\infty\}$ y $\{-\infty\}$, la cual denotaremos por $\mathfrak{L}\left(\overline{\mathbb{R}}\right)$, está formada por los conjuntos Lebesgue medibles en $\mathbb{R}$ y los conjuntos de la forma $B \cup \{\infty\}$, $B \cup \{-\infty\}$ y $B \cup \{-\infty, \infty\}$, donde $B \in \mathfrak{L}(\mathbb{R})$. Así que podemos extender la medida de Lebesgue a $\mathfrak{L}\left(\overline{\mathbb{R}}\right)$ definiendo $\lambda(\{-\infty, \infty\}) = 0$.

CAPÍTULO 3

FUNCIONES DE VARIACIÓN ACOTADA

Previamente al trabajo de Lebesgue, Thomas Joannes Stieltjes había extendido el concepto de integral en una dirección distinta a la de Lebesgue. En el año 1894 publicó un artículo titulado *Recherches sur les fractions continues* ([**87**]), donde planteó el problema de determinar el límite, si existe, de una fracción continua de la forma:

$$\frac{1}{a_1z+\frac{1}{a_2z+\frac{1}{a_3z+\frac{1}{\cdots}}}}$$

donde $(a_n)_{n\in\mathbb{N}}$ es una sucesión de números reales positivos y z un número real o un número complejo.

En el desarrollo que realizó Stieltjes en su artículo, obtuvo una expresión que lo llevó a introducir el concepto de momento de una función monótona creciente y al problema de la determinación de esa función a partir de sus momentos. Para ello, decía que una distribución de masa sobre la parte positiva de una recta de origen O representa una función creciente de la distancia x al origen. Agregaba que, inversamente, una función creciente, definida sobre la parte positiva de la recta, se puede imaginar como representando una distribución de masa. Dada una función creciente Φ, definida en un intervalo $[a,b]$ sobre la parte positiva de una recta, consideraba una partición $\{x_0,x_1,\ldots,x_n\}$ del intervalo $[a,b]$, tomaba un punto ξ_i en cada subintervalo $[x_{i-1},x_i]$, consideraba la suma $\sum_{i=1}^n \xi_i\left(\Phi\left(x_i\right)-\Phi\left(x_{i-1}\right)\right)$ y definía el momento de Φ como el límite de esa suma cuando las longitudes de los subintervalos de la partición tienden a cero (para $k\in\mathbb{N}$, el momento de orden k de Φ sería el límite de las sumas $\sum_{i=1}^n \xi_i^k\left(\Phi\left(x_i\right)-\Phi\left(x_{i-1}\right)\right)$). Generalizando esta idea, consideró una función continua f, definida sobre el intervalo $[a,b]$, y definió la integral de f con respecto a Φ en el intervalo $[a,b]$, denotada por $\int_a^b f\left(u\right)d\Phi\left(u\right)$, como el límite de las sumas $\sum_{i=1}^n f\left(\xi_i\right)\left(\Phi\left(x_i\right)-\Phi\left(x_{i-1}\right)\right)$. De esta forma surgió lo que ahora se conoce como la integral de Riemann-Stieltjes.

La integral de Stieltjes jugó un papel central para el desarrollo de una teoría general de la medida ya que permitió visualizar el problema de la integración de funciones en un contexto más amplio. Después que Lebesgue desarrolló su teoría de integración para funciones de $\mathbb{R}$ en $\mathbb{R}$, basándose en la teoría de la medida en $\mathbb{R}$, que él mismo desarrolló, la extensión al caso de funciones de $\mathbb{R}^n$ en $\mathbb{R}$, basada en el concepto de medida de Lebesgue en $\mathbb{R}^n$, fue relativamente simple. De hecho el mismo Lebesgue lo había hecho ya para el caso de $\mathbb{R}^2$. Sin embargo, la medida de Lebesgue en $\mathbb{R}$ se construye con base en el concepto de longitud de un intervalo; de ahí se puede pasar a definir la medida de un producto cartesiano de n intervalos $I_1 \times I_2 \times \cdots \times I_n$ como el producto de las longitudes de cada uno de esos intervalos. En cambio, la integral de Stieltjes hacía ver que el concepto de medida de un intervalo puede ser más amplio y no restringirse a definirla como la longitud de ese intervalo. Con el trabajo de Stieltjes pudo verse que la medida de un intervalo $[a, b]$ puede definirse también como la diferencia $\Phi(b) - \Phi(a)$, donde Φ es una función creciente (con las adecuaciones correspondientes para el caso en que Φ no es continua).

La integral de Stieltjes se vincula con otro concepto de importancia central, el de función de variación acotada. En su libro *Leçons sur l'intégration et la recherche des fonctions primitives* (1904), Lebesgue atribuye la invención de este concepto a Jordan. En ese libro, Lebesgue dedicó un capítulo al estudio de las funciones de variación acotada, aunque no motivado por el trabajo de Stieltjes, el cual ni siquiera menciona. La motivación de Lebesgue provenía del estudio que hacía de la rectificación de curvas, es decir de la medida de la longitud de una curva, y también del hecho de que, si $f : [a, b] \to \mathbb{R}$, es integrable, entonces la función $x \to \int_a^x f(y)\, dy$, definida sobre $[a, b]$, es de variación acotada.

La relación entre la integral de Stieltjes y las funciones de variación acotada proviene de que una función $g : [a, b] \to \mathbb{R}$ es de variación acotada si y sólo si se puede expresar como la diferencia de dos funciones no decrecientes, resultado que Lebesgue demostró en el libro citado. De aquí que la integral de Stieltjes se pueda extender al caso en que Φ es de variación acotada.

Además de los resultados y aplicaciones de las funciones de variación acotada que se hicieron previamente, el trabajo que las puso en el centro de la teoría de integración y que vinculó estrechamente la integral de Stieltjes con las funciones de variación acotada, fue un artículo de Frédéric Riesz del año 1909, titulado *Sur les opérations fonctionnelles linéaires* ([**80**]), en el cual demostró el ahora llamado teorema de representación de Riesz:

> Si A es una funcional lineal definida sobre el conjunto $\mathcal{C}$ de funciones continuas $f : [a, b] \to \mathbb{R}$ y si A es continua considerando la norma de la convergencia uniforme en $\mathcal{C}$, entonces existe una función de variación acotada $g : [a, b] \to \mathbb{R}$ tal que $A(f) = \int_a^b f dg$ para cualquier función $f \in \mathcal{C}$.

En lo que respecta a la Teoría de la Medida, la importancia de las funciones de variación acotada radica en el hecho de que cualquier medida (con signo, es decir que puede tomar valores positivos y negativos) sobre los conjuntos borelianos de $\mathbb{R}$ tal que los intervalos acotados tienen medida finita se puede considerar como generada por una función definida sobre $\mathbb{R}$ y que es de variación acotada sobre cualquier intervalo compacto. Esto lo desarrollaremos en el capítulo 4 de este libro.

En esta sección vamos a exponer la definición y las propiedades básicas de una función de variación acotada definida sobre un intervalo compacto. En la sección siguiente haremos un estudio más fino en lo que respecta a su parte continua y su parte de saltos.

DEFINICIÓN 3.1. *Dada una función* $g : [a, b] \to \mathbb{R}$ *y una partición* $P = \{x_0, x_1, \ldots, x_n\}$ *del intervalo* $[a, b]$*, definamos* $V_g(P) = \sum_{k=1}^{n} |g(x_k) - g(x_{k-1})|$*.*

Diremos que g *es de variación acotada en* $[a, b]$ *si:*

$V_g[a, b] = \sup \{V_g(P) : P$ *es una partición de* $[a, b]\} < \infty$*.*

Las funciones de variación acotada tienen algunas propiedades las cuales hacen que se pueda trabajar fácilmente con ellas. Una de esas propiedades es que toda función $g : \mathbb{R} \to \mathbb{R}$ tal que, restringida a cualquier intervalo compacto, es de de variación acotada, se puede expresar como la diferencia de dos funciones no decrecientes; de manera que, estudiando las propiedades de las funciones no decrecientes, podemos obtener las de las que, restringidas a un intervalo compacto, son de variación acotada. Una función no decreciente es fácil de tratar ya que el conjunto de sus discontinuidades es a lo más infinito numerable y todas sus discontinuidades son saltos. Además, las propiedades que nos van a interesar de una función no decreciente no se alteran si, en cada punto de discontinuidad, la sustituimos por su límite por la derecha o su límite por la izquierda; así que podemos asumir que la función es continua por la derecha para cualquier $x \in \mathbb{R}$, o bien continua por la izquierda para cualquier $x \in \mathbb{R}$. Si g es continua por la derecha, la función que asigna a cada intervalo de la forma $(a, b]$ el valor $g(b) - g(a)$ se puede extender a una medida definida sobre los conjuntos borelianos.De la misma manera, si g es continua por la izquierda, la función que asigna a cada intervalo de la forma $[a, b)$ el valor $g(b) - g(a)$ se puede extender a una medida definida sobre los conjuntos borelianos. De esta forma, hablar de funciones no decrecientes es equivalente a hablar de medidas definidas sobre los borelianos.

Pasemos entonces a demostrar los resultados que nos permiten probar las propiedades mencionadas. Como veremos, el teorema de Heine Borel juega un papel central en el desarrollo de la teoría.

PROPOSICIÓN 3.1. *Si* $g : [a, b] \to \mathbb{R}$ *es de variación acotada, entonces está acotada.*

Demostración

Dado cualquier punto $x \in [a, b]$, consideremos la partición $P = \{a, x, b\}$; se tiene entonces:

$|f(x)| - |f(a)| \leq |f(x) - f(a)| \leq |f(x) - f(a)| + |f(b) - f(x)| = V_g(P) \leq V_g[a, b]$.

Así que:

$|f(x)| \leq |f(a)| + V_g[a, b]$.

■

PROPOSICIÓN 3.2. *Sea* $g : [a, b] \to \mathbb{R}$ *y* $c \in [a, b]$*, entonces* $V_g[a, b] = V_g[a, c] + V_g[c, b]$*.*

Demostración

Si P_1 es una partición de $[a,c]$ y P_2 es una partición de $[c,b]$, entonces $P = P_1 \cup P_2$ es una partición de $[a,b]$ y se tiene:

$V_g(P_1) + V_g(P_2) = V_g(P) \leq V_g[a,b]$.

Por lo tanto:

$V_g[a,c] + V_g[c,b] \leq V_g[a,b]$.

Para probar la otra desigualdad, sea P una partición de $[a,b]$ y $P_0 = P \cup \{c\}$, entonces:

$P_1 = P_0 \cap [a,c]$ es una partición de $[a,c]$,

$P_2 = P_0 \cap [c,b]$ es una partición de $[c,b]$,

y se tiene:

$V_g(P_0) = V_g(P_1) + V_g(P_2) \leq V_g[a,c] + V_g[c,b]$.

Por otra parte, si $c \in P$, entonces $V_g(P) = V_g(P_0)$, mientras que si $c \notin P$, entonces $c \in (x_k, x_{k+1})$, donde x_k y x_{k+1} son dos puntos consecutivos de P, así que:

$|g(x_{k+1}) - g(x_k)| \leq |g(x_{k+1}) - g(c)| + |g\,(c) - g(x_k)|$.

Por lo tanto, en cualquier caso:

$V_g(P) \leq V_g(P_0) \leq V_g[a,c] + V_g[c,b]$.

Así que:

$V_g[a,b] \leq V_g[a,c] + V_g[c,b]$.

$\blacksquare$

Corolario 3.1. *Sea $g : [a,b] \to \mathbb{R}$ y $c \in [a,b]$, entonces g es de variación acotada en el intervalo $[a,b]$ si y sólo si es de variación acotada en cada uno de los intervalos $[a,c]$ y $[c,b]$.*

Corolario 3.2. *Sea $g : [a,b] \to \mathbb{R}$ y $[c,d] \subset [a,b]$, entonces $V_g\,[c,d] \leq V_g[a,b]$.*

Teorema 3.1. *El conjunto de funciones de variación acotada, definidas en un mismo intervalo $[a,b]$, forma un espacio vectorial.*

Demostración

Obviamente, si g es de variación acotada y $c \in \mathbb{R}$, entonces cg es de variación acotada.

Sean f y g funciones de variación acotada en el intervalo $[a,b]$ y $P = \{x_0, x_1, \ldots, x_n\}$ cualquier partición de $[a,b]$. Se tiene:

$V_{f+g}(P) = \sum_{i=1}^{n} |(f+g)(x_i) - (f+g)(x_{i-1})| = \sum_{i=1}^{n} |[f(x_i) - f(x_{i-1})] + [g(x_i) - g(x_{i-1})]|$

$\leq \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| + \sum_{i=1}^{n} |g(x_i) - g(x_{i-1})| \leq V_f[a,b] + V_g[a,b].$

Así que, tomando supremos:

$V_{f+g}[a,b] \leq V_f[a,b] + V_g[a,b] < \infty.$

Por lo tanto $f+g$ es de variación acotada en $[a,b]$.

■

PROPOSICIÓN 3.3. *Si $g : [a,b] \to \mathbb{R}$ es de variación acotada, entonces g^2 también lo es.*

Sea $M = \sup\{|g(x)| \mid x \in [a,b]\}$ y $P = \{x_0, x_1, \ldots, x_n\}$ cualquier partición del intervalo $[a,b]$, entonces:

$V_{g^2}(P) = \sum_{k=1}^{n} |g^2(x_k) - g^2(x_{k-1})| = \sum_{k=1}^{n} |g(x_k) + g(x_{k-1})|\,|g(x_k) - g(x_{k-1})|$

$\leq 2M \sum_{k=1}^{n} |g(x_k) - g(x_{k-1})| = 2MV_g(P) \leq 2MV_g[a,b].$

Por lo tanto:

$V_{g^2}[a,b] \leq 2MV_g[a,b] < \infty.$

Así que g^2 es de variación acotada.

■

COROLARIO 3.3. *Si $f : [a,b] \to \mathbb{R}$ y $g : [a,b] \to \mathbb{R}$ son funciones de variación acotada, entonces fg también lo es.*

Demostración

$fg = \frac{1}{2}\left[(f+g)^2 - f^2 - g^2\right].$

■

PROPOSICIÓN 3.4. *Si $g : [a,b] \to \mathbb{R}$ es una función monótona, entonces es de variación acotada.*

Demostración

Si $P = \{x_0, x_1, \ldots, x_n\}$ es cualquier partición de $[a,b]$, entonces:

$V_g(P) = \sum_{k=1}^{n} |g(x_k) - g(x_{k-1})| = |g(b) - g(a)|.$

Así que:

$V_g[a,b] = |g(b) - g(a)| < \infty.$

■

COROLARIO 3.4. *Si $g_1 : [a,b] \to \mathbb{R}$ y $g_2 : [a,b] \to \mathbb{R}$ son funciones no decrecientes, entonces $g_1 - g_2$ es de variación acotada.*

COROLARIO 3.5. *Si* $g_1 : \mathbb{R} \to \mathbb{R}$ *y* $g_2 : \mathbb{R} \to \mathbb{R}$ *son funciones no decrecientes, entonces* $g_1 - g_2$ *es de variación acotada sobre cualquier intervalo compacto.*

TEOREMA 3.2. *Si* $g : \mathbb{R} \to \mathbb{R}$ *es una función de variación acotada sobre cualquier intervalo compacto, entonces se puede expresar como la diferencia de dos funciones no decrecientes.*

Demostración

La idea de la demostración consiste en definir una función V, no decreciente, tal que $V - g$ también sea no decreciente.

Hay un problemita que es necesario salvar para poder definir la función V; se trata de que estamos tomando una función g definida sobre todo $\mathbb{R}$ de la cual únicamente sabemos que es de variación acotada sobre cualquier intervalo compacto.

Si $a \in \mathbb{R}$, definamos $V : [a, \infty) \to \mathbb{R}$ como $V(x) = V_g[a, x]$ y tomemos $x, y \in [a, \infty)$ tales que $x < y$; entonces se tiene:

$V_g[a, y] = V_g[a, x] + V_g[x, y]$

Así que:

$V(y) - V(x) = V_g[x, y] \geq 0$

Por lo tanto, V es una función no decreciente.

Además, como $V_g[x, y] \geq g(y) - .g(x)$, entonces $V(y) - V(x) \geq g(y) - .g(x)$; por lo tanto:

$V(y) - g(y) \geq V(x) - .g(x)$

Así que $V - g$ es no decreciente sobre el intervalo $[a, \infty)$.

Pero, como decíamos, g está definida sobre todo $\mathbb{R}$; así que es necesario hacer algunos ajustes en la definición de V con el objeto de tenerla definida sobre todo $\mathbb{R}$. Para esto, podemos tomar un número real arbitrario a_0 y primero definir V sobre el intervalo $(-\infty, a_0]$ y después sobre el intervalo $[a_0, \infty)$.

Sobre $(-\infty, a_0]$ podemos definir $V(x) = -V_g[x, a_0]$ (el signo $-$ es para que V sea no decreciente); mientras que sobre $[a_0, \infty)$ podemos definir $V(x) = V_g[a_0, x]$. La función V así definida toma valores menores o iguales a cero sobre $(-\infty, a_0]$ y valores mayores o iguales a cero sobre $[a_0, \infty)$ (en 0 toma el valor 0); así que es una función no decreciente definida sobre todo $\mathbb{R}$.

Por comodidad vamos a tomar $a_0 = 0$.

Definamos entonces la función $V : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:

$$V(x) = \begin{cases} -V_g[x,0] & \text{si } x \in (-\infty, 0) \\ 0 & \text{si } x = 0 \\ V_g[0,x] & \text{si } x \in (0, \infty) \end{cases}$$

Si $x < y$, se tiene:

$V(y) = V(x) + V_g[x, y]$.

Así que:

$V(y) - V(x) = V_g[x, y] \geq 0$.

Por lo tanto:

$V(y) \geq V(x)$.

Así que V es una función no decreciente.

Además:

$V_g[x, y] \geq |g(y) - g(x)| \geq g(y) - g(x)$.

Por lo tanto:

$V(y) - V(x) \geq g(y) - g(x)$.

Así que:

$V(y) - g(y) \geq V(x) - g(x)$.

Es decir, $V - g$ es una función no decreciente.

Definiendo $h_1 = V$ y $h_2 = V - g$, se tiene:

$g = h_1 - h_2$.

∎

Combinando los dos últimos resultados, se tiene el siguiente:

TEOREMA 3.3. *Una función $g : \mathbb{R} \to \mathbb{R}$ es de variación acotada sobre cualquier intervalo compacto si y sólo si se puede expresar como la diferencia de dos funciones no decrecientes.*

3.1. Estudio de las discontinuidades de una función de variación acotada

A continuación vamos a estudiar las funciones $g : \mathbb{R} \to \mathbb{R}$ que son de variación acotada sobre cualquier intervalo compacto. En particular vamos a analizar el conjunto de discontinuidades de una función de ese tipo. Como toda función de variación acotada sobre cualquier intervalo compacto se puede expresar como la diferencia de dos funciones no decrecientes, trataremos primero el caso de una función no decreciente.

Primero veremos algunas propiedades elementales:

Sea $f : \mathbb{R} \to \mathbb{R}$ una función no decreciente. Por ser monótona, f tiene límites por la derecha y por la izquierda en todo punto $x \in \mathbb{R}$; en efecto, sea $x_0 \in \mathbb{R}$, entonces:

$f(x) \geq f(x_0)$ para cualquier $x > x_0$.

$f(x) \leq f(x_0)$ para cualquier $x < x_0$.

Así que el conjunto $\{f(x) : x > x_0\}$ está acotado por abajo y el conjunto $\{f(x) : x < x_0\}$ está acotado por arriba. Por lo tanto:

$f(x_0+) = \text{lím}_{x \to x_0+} f(x) = \text{ínf}\{f(x) : x > x_0\} \in \mathbb{R}$

$f(x_0-) = \text{lím}_{x \to x_0-} f(x) = \sup\{f(x) : x < x_0\} \in \mathbb{R}$

Por lo anterior, f no tiene discontinuidades oscilatorias; es decir, todas sus discontinuidades son de salto.

Además, si a y b son dos números reales tales que $a < b$, entonces:

1. Si $t \in (a, b)$, f es discontinua en t si y sólo si $f(t+) > f(t-)$; en ese caso, f tiene un salto en t de magnitud $f(t+) - f(t-)$.

2. Si $t \in [a, b)$, f es discontinua por la derecha en t si y sólo si $f(t+) > f(t)$; en ese caso, f tiene un salto por la derecha en t de magnitud $f(t+) - f(t)$.

3.Si $t \in (a, b]$, f es discontinua por la izquierda en t si y sólo si $f(t) > f(t-)$; en ese caso, f tiene un salto por la izquierda en t de magnitud $f(t) - f(t-)$.

Por otra parte:

$i)$ La función $f^i : \mathbb{R} \to \mathbb{R}$ definida por $f^i(x) = f(x-)$ es no decreciente y continua por la izquierda. Además, $f^i(x+) = f(x+)$ y $f^i(x+) - f^i(x) = f(x+) - f(x-)$ para cualquier $x \in \mathbb{R}$.

$ii)$ La función $f^d : \mathbb{R} \to \mathbb{R}$ definida por $f^d(x) = f(x+)$ es no decreciente y continua por la derecha. Además, $f^d(x-) = f(x-)$ y $f^d(x) - f^d(x-) = f(x+) - f(x-)$ para cualquier $x \in \mathbb{R}$.

En efecto:
$i)$ Si $x < y$, $f(x-) \leq f(y-)$, así que f^i es no decreciente.

Dado $x \in \mathbb{R}$ y $\varepsilon > 0$, existe $\delta > 0$ tal que $f(x-) - f(y) < \varepsilon$ para cualquier $y \in (x - \delta, x)$.

Dado $z \in (x-\delta, x)$, tomemos $y \in (x-\delta, z)$, entonces $y \in (x-\delta, x)$ y $f(y) \leq f(z-)$, así que:

$f(x-) - f(z-) \leq f(x-) - f(y) < \varepsilon$.

Por lo tanto f^i es continua por la izquierda.

Para cualquier $x \in \mathbb{R}$ se tiene $f(x) \geq f(x-)$, es decir $f(x) \geq f^i(x)$, así que $f(x+) \geq f^i(x+)$.

Dado $x \in \mathbb{R}$ y $\varepsilon > 0$, tomemos $\delta > 0$ tal que $f^i(y) - f^i(x+) < \varepsilon$ para cualquier $y \in (x, x+\delta)$.

Si $y \in (x, x+\delta)$, entonces $y > x$, así que $f(y-) \geq f(x+)$, es decir, $f^i(y) \geq f(x+)$, por lo tanto:

$0 \leq f(x+) - f^i(x+) \leq f^i(y) - f^i(x+) < \varepsilon$.

Como esto es válido para cualquier $\varepsilon > 0$, se tiene $f^i(x+) = f(x+)$.

Por lo tanto, $f^i(x+) - f^i(x) = f(x+) - f(x-)$ para cualquier $x \in \mathbb{R}$, de lo cual se sigue que $f^i(x+) = f(x+)$.

El inciso ii se demuestra de manera similar.
Además, tenemos el siguiente resultado:

PROPOSICIÓN 3.5. *Sea* $f : \mathbb{R} \to \mathbb{R}$ *una función no decreciente. Si* $a < b$, $k \in \mathbb{N}$, $x_1, x_2, \ldots, x_k \in (a,b)$ *y* $x_1 < x_2 < \cdots < x_k$, *entonces:*

$$f(b) - f(a) \geq f(a+) - f(a) + \sum_{j=1}^{k} [f(x_j+) - f(x_j-)] + f(b) - f(b-)$$

Demostración

Este resultado es fácil de visualizar, pero vamos a demostrarlo analiticamente:

Definamos $x_0 = a$ y $x_{k+1} = b$ y, para cada $j \in \{0, 1, 2, \ldots, k\}$, sea $y_j \in (x_j, x_{j+1})$; entonces:

$$f(a+) - f(a) + \sum_{j=1}^{k} [f(x_j+) - f(x_j-)] + f(b) - f(b-)$$

$$\leq f(a+) - f(a) + \sum_{j=1}^{k} [f(y_j) - f(y_{j-1})] + f(b) - f(b-)$$

$$= f(a+) - f(a) + f(y_k) - f(y_0) + f(b) - f(b-)$$

$$\leq f(a+) - f(a) + f(b-) - f(a+) + f(b) - f(b-)$$

$$= f(b) - f(a)$$

∎

Proposición 3.6. *Toda función no decreciente $f : \mathbb{R} \to \mathbb{R}$ tiene a lo más un conjunto numerable de discontinuidades.*

Demostración

Sean $m, n \in \mathbb{N}$ y definamos:

$A_{m,n} = \left\{x \in (-m, m) : f(x+) - f(x-) > \frac{1}{n}\right\}$

$A_m = \{x \in (-m, m) : f(x+) \neq f(x-)\}$

$A = \{x \in \mathbb{R} : f(x+) \neq f(x-)\}$

Sea $M \in N$ tal que $\frac{M}{n} > f(m) - f(-m)$.

Si $k \in \mathbb{N}$, $x_1, x_2, \ldots, x_k \in A_{m,n}$ y $x_1 < x_2 < \cdots < x_k$, entonces:

$f(m) - f(-m) \geq \sum_{j=1}^{k} [f(x_k+) - f(x_k-)]$

Así que:

$\frac{k}{n} < \sum_{j=1}^{k} [f(x_j+) - f(x_j-)] < \frac{M}{n}$

Por lo tanto, $k < M$; así que $A_{m,n}$ es un conjunto finito.

Además:

$A_m = \bigcup_{n=1}^{\infty} A_{m,n}$

Por lo tanto, A_m es un conjunto a lo más infinito numerable.

Finalmente:

$A = \bigcup_{m=1}^{\infty} A_m$

Así que, también A es un conjunto a lo más infinito numerable.

■

Corolario 3.6. *Una función de variación acotada sobre cualquier intervalo compacto tiene a lo más un conjunto numerable de discontinuidades.*

Ahora podemos extender la proposición 3.5, abarcando todas las discontinuidades de una función no decreciente.

Proposición 3.7. *Sea $f : \mathbb{R} \to \mathbb{R}$ una función no decreciente, a y b dos números reales tales que $a < b$ y $D_{(a,b)}$ el conjunto de puntos del intervalo (a, b) donde f es discontinua; entonces:*

$f(b) - f(a) \geq f(a+) - f(a) + \sum_{\{y \in D_{(a,b)}\}} [f(y+) - f(y-)] + f(b) - f(b-)$

Demostración

Si $D_{(a,b)}$ es vacío, el resultado es trivial, mientras que si $D_{(a,b)}$ es un conjunto no vacío finito, el resultado ya se demostró previamente. Así que únicamente resta probarlo para el caso en que $D_{(a,b)}$ es un conjunto infinito numerable.

Sea $D_{(a,b)} = \{y_1, y_2, \ldots\}$

Por la proposición 3.5, para cada $n \in \mathbb{N}$, se tiene:

$$f(b) - f(a) \geq f(a+) - f(a) + \sum_{j=1}^{n} [f(y_j+) - f(y_j-)] + f(b) - f(b-)$$

Tomando límites, obtenemos:

$$f(b) - fa \geq f(a+) - f(a) + \sum_{\{y \in D_{(a,b)}\}} [f(y+) - f(y-)] + f(b) - f(b-)$$

■

Corolario 3.7. *Sea $f : \mathbb{R} \to \mathbb{R}$ una función no decreciente, a y b dos números reales tales que $a < b$ y $D^d_{[a,b)}$ el conjunto de puntos del intervalo $[a, b)$ donde f es discontinua por la derecha; entonces:*

$$f(b) - f(a) \geq \sum_{\{y \in D^d_{[a,b)}\}} [f(y+) - f(y)] + f(b) - f(b-)$$

Lema 3.1. ***Demostración***

$$f(b) - f(a) \geq f(a+) - f(a) + \sum_{\{y \in D_{(a,b)}\}} [f(y+) - f(y-)] + f(b) - f(b-)$$

$$\geq f(a+) - f(a) + \sum_{\{y \in D_{(a,b)}\}} [f(y+) - f(y)] + f(b) - f(b-)$$

$$= \sum_{\{y \in D^d_{[a,b)}\}} [f(y+) - f(y)] + f(b) - f(b-)$$

■

Corolario 3.8. *Sea $f : \mathbb{R} \to \mathbb{R}$ una función no decreciente, a y b dos números reales tales que $a < b$ y $D^i_{(a,b]}$ el conjunto de puntos del intervalo $(a, b]$ donde f es discontinua por la izquierda; entonces:*

$$f(b) - f(a) \geq f(a+) - f(a) + \sum_{\{y \in D^i_{(a,b]}\}} [f(y) - f(y-)]$$

Demostración

$$f(b) - f(a) \geq f(a+) - f(a) + \sum_{\{y \in D_{(a,b)}\}} [f(y+) - f(y-)] + f(b) - f(b-)$$

$$\geq f(a+) - f(a) + \sum_{\{y \in D_{(a,b)}\}} [f(y) - f(y-)] + f(b) - f(b-)$$

$$= f(a+) - f(a) + \sum_{\{y \in D^i_{(a,b]}\}} [f(y) - f(y-)]$$

■

Ahora vamos a afinar el resultado del teorema 3.2. Queremos demostrar que si $g : \mathbb{R} \to \mathbb{R}$ es una función de variación acotada sobre cualquier intervalo compacto, entonces se puede expresar como la diferencia de dos funciones no decrecientes, $f_1 : \mathbb{R} \to \mathbb{R}$ y $f_2 : \mathbb{R} \to \mathbb{R}$, las cuales no tienen discontinuidades en común del mismo lado.La idea es que al tomar la diferencia de dos funciones no decrecientes, si éstas tienen una discontinuidad por la derecha en un mismo punto t, entonces, si le quitamos a ambas, en ese punto, la magnitud del más pequeño de los dos saltos por la derecha, una de las dos funciones que se obtienen es continua por la derecha en ese punto. De la misma manera, si éstas tienen una discontinuidad por la izquierda en un mismo punto t, entonces, si le quitamos a ambas, en ese punto, la magnitud del más pequeño de los dos saltos por la izquierda, una de las dos funciones que se obtienen es continua por la izquierda en ese punto. De esta forma, las nuevas funciones no tienen discontinuidades en común del mismo lado y la diferencia entre ellas es igual a la diferencia entre las dos funciones originales.

Como vamos a tener que tratar con el conjunto de discontinuidades de una función no decreciente, el cual puede ser infinito numerable, vamos a requerir de algunos resultados que nos permitan tratar con toda generalidad el problema mencionado.

El siguiente lema es un resultado de Teoría de la Medida, pero como ese tema no lo hemos tratado, vamos a hacer la demostración de manera directa.

Lema 3.2. *Sea $\mathbb{F} = \{x_1, x_2, \ldots\}$ un conjunto infinito numerable y $(z_n)_{n\in\mathbb{N}}$ una sucesión de números reales no negativos. Para cada $n \in \mathbb{N}$, definamos $m(x_n) = z_n$ y, para cualquier subconjunto A de $\mathbb{F}$, definamos:*

$\mu(A) = \sum_{\{x\in A\}} m(x)$.

Entonces μ es una función no negativa y σ-aditiva, con valores en $\overline{\mathbb{R}}$, definida sobre la σ-álgebra formada por todos los subconjuntos de $\mathbb{F}$.

Demostración

Obviamente μ es no negativa y $\mu(\emptyset) = 0$.

Agreguemos a Ω un elemento arbitrario, el cual denotaremos por x_0, y definamos $m(x_0) = 0$.

Sea $A_1, A_2, \ldots$ una colección infinita numerable de subconjuntos de Ω tales que $A_i \cap A_j = \emptyset$ para $i \neq j$. Denotemos por $x_{s_1}, x_{s_2}, \ldots$ a los elementos de la unión $\bigcup_{i=1}^{\infty} A_i$ y para cada $i \in \mathbb{N}$, denotemos por $x_{i,1}, x_{i,2}, \ldots$ a los elementos de A_i. Si A es vacío o un conjunto finito y r es el número de sus elementos, definamos $x_{s_k} = x_0$ para cualquier $k \in \{r+1, r+2, \ldots\}$ y, para cada $i \in \mathbb{N}$, si A_i es vacío o un conjunto finito y r_i es el número de sus elementos, definamos $x_{i,j} = x_0$ para cualquier $j \in \{r_i+1, r_i+2, \ldots\}$.

Para cada terna $i, j, k \in \mathbb{N}$, definamos:

$m_{ij} = m\left(x_{i,j}\right)$,

$m_k = m\left(x_{s_k}\right)$.

Entonces:

$\mu\left(A_i\right) = \sum_{j=1}^{\infty} m_{ij}$,

$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{k=1}^{\infty} m_k$.

Por otra parte, la sucesión $\left(m_k\right)_{k\in\mathbb{N}}$ contiene a todos los elementos m_{ij} tales que $x_{i,j} \in A_i$; así que, para cualquier pareja $n, m \in \mathbb{N}$, se tiene:

$\sum_{i=1}^{n}\left(\sum_{j=1}^{m} x_{ij}\right) \leq \sum_{k=1}^{\infty} m_k$.

Así que:

$\sum_{1=1}^{\infty} \mu\left(A_i\right) = \sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty} m_{ij}\right) \leq \sum_{k=1}^{\infty} m_k = \mu\left(\bigcup_{i=1}^{\infty} A_i\right)$.

Además, la sucesión doble $\left(m_{ij}\right)_{i,j\in\mathbb{N}}$ contiene a todos los elementos m_k tales que

$x_{s_k} \in \bigcup_{i=1}^{\infty} A_i$;

en consecuencia, dado $N \in \mathbb{N}$, existen $n, m \in \mathbb{N}$ tales que:

$\sum_{k=1}^{N} m_k \leq \sum_{i=1}^{n}\left(\sum_{j=1}^{m} m_{ij}\right)$.

Así que:

$\sum_{k=1}^{N} m_k \leq \sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty} m_{ij}\right) = \sum_{1=1}^{\infty} \mu\left(A_i\right)$.

Por lo tanto:

$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{k=1}^{\infty} m_k \leq \sum_{1=1}^{\infty} \mu\left(A_i\right)$.

Podemos concluir entonces que:

$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{1=1}^{\infty} \mu\left(A_i\right)$.

Esta relación incluye la propiedad de la aditividad finita ya que, si $A_1, \ldots, A_n$ es una colección finita de subconjuntos de $\mathbb{F}$, ajenos por parejas, podemos definir $A_k = \emptyset$ para cualquier $k \in \{n+1, n+2, \ldots\}$. Entonces, $\mu\left(A_k\right) = 0$ para cualquier $k \in \{n+1, n+2, \ldots\}$, así que:

$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{1=1}^{\infty} \mu\left(A_i\right) = \sum_{1=1}^{n} \mu\left(A_i\right)$.

$\blacksquare$

COROLARIO 3.9. *Sea $\{x_1, x_2, \ldots\}$ un conjunto infinito numerable de números reales y $\left(z_n\right)_{n\in\mathbb{N}}$ una sucesión de números reales no negativos. Para cada subconjunto A de $\mathbb{R}$, definamos:*

$\mu(A) = \sum_{\{j\in\mathbb{N}:x_j\in A\}} z_j.$

Entonces μ es una función no negativa y σ-aditiva, con valores en $\overline{\mathbb{R}}$, definida sobre la σ-álgebra formada por todos los subconjuntos de $\mathbb{R}$.

Los siguientes dos resultados se siguen inmediatamente de la proposición 2.1.

Ejercicio 3.1. *Sea $(q_n)_{n\in\mathbb{N}}$ una sucesión de números reales no negativos tales que la serie $\sum_{n=1}^{\infty} q_n$ converge. Si $(A_n)_{n\in\mathbb{N}}$ es una sucesión creciente de subconjuntos de $\mathbb{N}$, entonces:*

$\sum_{\{j\in\cup_{n=1}^{\infty}A_n\}} q_j = \lim_{n\rightsquigarrow\infty} \sum_{\{j\in A_n\}} q_j.$

Ejercicio 3.2. *Sea $(q_n)_{n\in\mathbb{N}}$ una sucesión de números reales no negativos tales que la serie $\sum_{n=1}^{\infty} q_n$ converge. Si $(A_n)_{n\in\mathbb{N}}$ es una sucesión decreciente de subconjuntos de $\mathbb{N}$, entonces:*

$\sum_{\{j\in\cap_{n=1}^{\infty}A_n\}} q_j = \lim_{n\rightsquigarrow\infty} \sum_{\{j\in A_n\}} q_j.$

Teorema 3.4. *Si $g:\mathbb{R}\to\mathbb{R}$ es una función de variación acotada sobre cualquier intervalo compacto, entonces se puede expresar como la diferencia de dos funciones no decrecientes, $f_1:\mathbb{R}\to\mathbb{R}$ y $f_2:\mathbb{R}\to\mathbb{R}$, las cuales no tienen discontinuidades en común del mismo lado.*

Demostración

Definamos la función $V:\mathbb{R}\to\mathbb{R}$ como en la proposición 3.2, entonces $h_1 = V$ y $h_2 = V - g$ son no decrecientes y $g = h_1 - h_2$.

Sea $D = \{d_1, d_2, \ldots\}$ el conjunto de puntos donde tanto h_1 como h_2 son discontinuas y definamos:

$f_1(x) = h_1(x) - \sum_{\{z\in D:z<x\}} \min(h_2(z+) - h_2(z), h_1(z+) - h_1(z)) - \sum_{\{z\in D:z\leq x\}} \min(h_2(z) - h_2(z-), h_1($

$f_2(x) = h_2(x) - \sum_{\{z\in D:z<x\}} \min(h_2(z+) - h_2(z), h_1(z+) - h_1(z)) - \sum_{\{z\in D:z\leq x\}} \min(h_2(z) - h_2(z-), h_1($

Entonces:

$f_1 - f_2 = h_1 - h_2 = g.$

Si $x < y$, se tiene:

$f_1(y) - f_1(x) = h_1(y) - h_1(x)$

$- \sum_{\{z\in D:x\leq z<y\}} \min(h_2(z+) - h_2(z), h_1(z+) - h_1(z)) - \sum_{\{z\in D:x<z\leq y\}} \min(h_2(z) - h_2(z-), h_1(z) - h_1(z$

$\geq h_1(y) - h_1(x) - \sum_{\{z\in D:x\leq z<y\}} [h_1(z+) - h_1(z)] - \sum_{\{z\in D:x<z\leq y\}} [h_1(z) - h_1(z-)]$

$= h_1(y) - h_1(x) - \sum_{\{z\in D:x<z<y\}} [h_1(z+) - h_1(z-)] - h_1(x+) + h_1(x) - h_1(y) + h_1(y-).$

Pero, por el lema **??**:

$h_1(y) - h_1(x) - \sum_{\{z\in D: x<z<y\}} [h_1(z+) - h_1(z-)] \geq h_1(x+) - h_1(x) + h_1(y) - h_1(y-)$.

Por lo tanto:

$f_1(y) - f_1(x) \geq 0$.

De la misma manera, $f_2(y) - f_2(x) \geq 0$.

Así que tanto f_1 como f_2 son no decrecientes.

Mostremos ahora que f_1 y f_2 no tienen discontinuidades en común del mismo lado.

Sea $x \in D$ y $(x_n)_{n\in\mathbb{N}}$ una sucesión decreciente que tienda a x, entonces, por el corolario 3.2:

$f_1(x+) = \text{lím}_{n\rightsquigarrow\infty} f_1(x_n) = h_1(x+)$

$- \text{lím}_{n\to\infty} \sum_{\{z\in D: z<x_n\}} \text{mín}(h_2(z+) - h_2(z), h_1(z+) - h_1(z)) - \text{lím}_{n\rightsquigarrow\infty} \sum_{\{z\in D: z\leq x_n\}} \text{mín}(h_2(z) - h_2(z-),$

$= h_1(x+) - \sum_{\bigcap_{n=1}^{\infty}\{z\in D: z<x_n\}} \text{mín}(h_2(z+) - h_2(z), h_1(z+) - h_1(z)) - \sum_{\bigcap_{n=1}^{\infty}\{z\in D: z\leq x_n\}} \text{mín}(h_2(z) - h_2(z$

$= h_1(x+) - \sum_{\{z\in D: z\leq x\}} \text{mín}(h_2(z+) - h_2(z), h_1(z+) - h_1(z)) - \sum_{\{z\in D: z\leq x\}} \text{mín}(h_2(z) - h_2(z-), h_1(z) -$

Así que:

$f_1(x+) - f_1(x) = h_1(x+) - h_1(x) - \text{mín}(h_2(x+) - h_2(x), h_1(x+) - h_1(x))$

De la misma manera:

$f_2(x+) - f_2(x) = h_2(x+) - h_2(x) - \text{mín}(h_2(x+) - h_2(x), h_1(x+) - h_1(x))$

Por lo tanto:

$$f_1(x+) - f_1(x) = \begin{cases} 0 & \text{si } h_1(x+) - h_1(x) \leq h_2(x+) - h_2(x) \\ h_1(x+) - h_1(x) - (h_2(x+) - h_2(x)) & \text{si } h_1(x+) - h_1(x) > h_2(x+) - h_2(x) \end{cases}$$

$$f_2(x+) - f_2(x) = \begin{cases} h_2(x+) - h_2(x) - (h_1(x+) - h_1(x)) & \text{si } h_1(x+) - h_1(x) \leq h_2(x+) - h_2(x) \\ 0 & \text{si } h_1(x+) - h_1(x) > h_2(x+) - h_2(x) \end{cases}$$

Así que, f_1 y f_2 no tienen discontinuidades en común por la derecha.

Sea $x \in D$ y $(x_n)_{n\in\mathbb{N}}$ una sucesión creciente que tienda a x, entonces, por el corolario 3.1:

$f_1(x-) = \text{lím}_{n\rightsquigarrow\infty} f_1(x_n) = h_1(x-)$.

$- \text{lím}_{n\to\infty} \sum_{\{z\in D: z<x_n\}} \text{mín}(h_2(z+) - h_2(z), h_1(z+) - h_1(z)) - \text{lím}_{n\rightsquigarrow\infty} \sum_{\{z\in D: z\leq x_n\}} \text{mín}(h_2(z) - h_2(z-),$

$= h_1(x-) - \sum_{\bigcup_{n=1}^{\infty}\{z\in D: z<x_n\}} \text{mín}(h_2(z+) - h_2(z), h_1(z+) - h_1(z)) - \sum_{\bigcup_{n=1}^{\infty}\{z\in D: z\leq x_n\}} \text{mín}(h_2(z) - h_2(z$

$= h_1(x-) - \sum_{\{z\in D: z<x\}} \text{mín}(h_2(z+) - h_2(z), h_1(z+) - h_1(z)) - \sum_{\{z\in D: z<x\}} \text{mín}(h_2(z) - h_2(z-), h_1(z) -$

Así que:

$$f_1(x) - f_1(x-) = h_1(x) - h_1(x-) - \min(h_2(z) - h_2(z-), h_1(z) - h_1(z-))$$

De la misma manera:

$$f_2(x) - f_2(x-) = h_2(x) - h_2(x-) - \min(h_2(z) - h_2(z-), h_1(z) - h_1(z-))$$

Por lo tanto:

$$f_1(x) - f_1(x-) = \begin{cases} 0 & \text{si } h_1(z) - h_1(z-) \leq h_2(z) - h_2(z-) \\ h_1(x) - h_1(x-) - (h_2(z) - h_2(z-)) & \text{si } h_1(z) - h_1(z-) > h_2(z) - h_2(z-) \end{cases}$$

$$f_2(x+) - f_2(x) = \begin{cases} h_2(x) - h_2(x-) - (h_1(z) - h_1(z-)) & \text{si } h_1(z) - h_1(z-) \leq h_2(z) - h_2(z-) \\ 0 & \text{si } h_1(z) - h_1(z-) > h_2(z) - h_2(z-) \end{cases}$$

Por lo tanto, f_1 y f_2 no tienen discontinuidades en común por la izquierda.

Obsérvese que, si f es una función no decreciente y continua, entonces $f_1 + f$ y $f_2 + f$ satisfacen también el enunciado de la proposición.

■

Corolario 3.10. *Si $g : \mathbb{R} \to \mathbb{R}$ es una función continua por la derecha y de variación acotada sobre cualquier intervalo compacto, entonces se puede expresar como la diferencia de dos funciones no decrecientes continuas por la derecha.*

Corolario 3.11. *Si $g : \mathbb{R} \to \mathbb{R}$ es una función continua por la izquierda y de variación acotada sobre cualquier intervalo compacto, entonces se puede expresar como la diferencia de dos funciones no decrecientes continuas por la izquierda.*

Corolario 3.12. *Si $g : \mathbb{R} \to \mathbb{R}$ es una función continua y de variación acotada sobre cualquier intervalo compacto, entonces se puede expresar como la diferencia de dos funciones no decrecientes continuas.*

3.2. Parte continua y parte de saltos de una función de variación acotada

Como el conjunto de discontinuidades de una función no decreciente es a lo más infinito numerable, la función se puede descomponer en una función no decreciente que crece únicamente mediante saltos y una función no decreciente continua. Más adelante demostraremos este resultado. Primero vamos a demostrar que una función no decreciente se puede expresar como la suma de función no decreciente, continua por la derecha y que crece únicamente mediante saltos y una función no decreciente continua por la izquierda.

Teorema 3.5. *Sea $f : \mathbb{R} \to \mathbb{R}$ una función no decreciente, $D = \{d_1, d_2, \ldots\}$ el conjunto de puntos donde f es discontinua y definamos $f^d : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:*

$$f^d(x) = \begin{cases} f(0-) - \sum_{\{y \in D : x < y < 0\}} [f(y) - f(y-)] & si\ x \in (-\infty, 0) \\ f(0) + \sum_{\{y \in D : 0 < y \leq x\}} [f(y) - f(y-)] & si\ x \in [0, \infty) \end{cases}$$

Entonces f^d *es no decreciente, continua por la derecha, crece únicamente mediante saltos y* $f^d(x) - f^d(x-) = f(x) - f(x-)$ *para cualquier* $x \in \mathbb{R}$.

Además, la función $h = f - f^d$ *es no decreciente, continua por la izquierda y* $h(x+) - h(x) = f(x+) - f(x)$ *para cualquier* $x \in \mathbb{R}$.

Demostración

Si $x < z < 0$, se tiene:

$f^d(z) - f^d(x) = \sum_{\{y \in D : x < y < 0\}} [f(y) - f(y-)] - \sum_{\{y \in D : z < y < 0\}} [f(y) - f(y-)]$

$= \sum_{\{y \in D : x < y \leq z\}} [f(y) - f(y-)]$.

Si $0 \leq x < z$, se tiene:

$f^d(z) - f^d(x) = \sum_{\{y \in D : 0 < y \leq z\}} [f(y) - f(y-)] - \sum_{\{y \in D : 0 < y \leq x\}} [f(y) - f(y-)]$

$= \sum_{\{y \in D : x < y \leq z\}} [f(y) - f(y-)]$.

Si $x < 0$

$f^d(0) - f^d(x) = f(0) - f(0-) + \sum_{\{y \in D : x < y < 0\}} [f(y) - f(y-)]$.

Así que, para cualquier parera (x, z) de números reales tales que $x < z$ se tiene:

$f^d(z) - f^d(x) = \sum_{\{y \in D : x < y \leq z\}} [f(y) - f(y-)] \geq 0$

Así que f^d es una función no decreciente.

Sea $x \in \mathbb{R}$ y $(x_n)_{n \in \mathbb{N}}$ una sucesión decreciente que tienda a x.

Si $x \geq 0$, se tiene:

$f^d(x+) = \text{lím}_{n \rightsquigarrow \infty} f^d(x_n) = f(0) + \text{lím}_{n \rightsquigarrow \infty} \sum_{\{y \in D : 0 < y \leq x_n\}} [f(y) - f(y-)]$

$= f(0) + \sum_{\bigcap_{n=1}^{\infty} \{y \in D : 0 < y \leq x_n\}} [f(y) - f(y-)] = f(0) + \sum_{\{y \in D : 0 < y \leq x\}} [f(y) - f(y-)] = f^d(x)$.

Si $x < 0$, se tiene:

$f^d(x+) = \text{lím}_{n \rightsquigarrow \infty} f^d(x_n) = f(0) - \text{lím}_{n \rightsquigarrow \infty} \sum_{\{y \in D : x_n < y < 0\}} [f(y) - f(y-)]$

$= f(0) - \sum_{\bigcup_{n=1}^{\infty} \{y \in D : x_n < y < 0\}} [f(y) - f(y-)] = f(0) - \sum_{\{y \in D : x < y < 0\}} [f(y) - f(y-)] = f^d(x)$.

Por lo tanto, f^d es continua por la derecha.

Sea $x \in \mathbb{R}$ y $(x_n)_{n\in\mathbb{N}}$ una sucesión creciente que tienda a x.

Si $x \geq 0$, se tiene:

$f^d(x-) = \text{lím}_{n\rightsquigarrow\infty} f^d(x_n) = f(0) + \text{lím}_{n\rightsquigarrow\infty} \sum_{\{y\in D:0<y\leq x_n\}} [f(y) - f(y-)]$

$= f(0) + \sum_{\bigcup_{n=1}^{\infty}\{y\in D:0<y\leq x_n\}} [f(y) - f(y-)] = f(0) + \sum_{\{y\in D:0<y<x\}} [f(y) - f(y-)]$.

$f^d(x) - f^d(x-) = \sum_{\{y\in D:0<y\leq x\}} [f(y) - f(y-)] - \sum_{\{y\in D:0<y<x\}} [f(y) - f(y-)]$

$= f(x) - f(x-)$.

Si $x < 0$, se tiene:

$f^d(x-) = \text{lím}_{n\rightsquigarrow\infty} f^d(x_n) = f(0) - \text{lím}_{n\rightsquigarrow\infty} \sum_{\{y\in D:x_n<y<0\}} [f(y) - f(y-)]$

$= f(0) - \sum_{\bigcap_{n=1}^{\infty}\{y\in D:x_n<y<0\}} [f(y) - f(y-)] = f(0) - \sum_{\{y\in D:x\leq y<0\}} [f(y) - f(y-)]$.

$f^d(x) - f^d(x-) = \sum_{\{y\in D:x\leq y<0\}} [f(y) - f(y-)] - \sum_{\{y\in D:x<y<0\}} [f(y) - f(y-)]$

$= f(x) - f(x-)$.

Así que el conjunto de discontinuidades por la izquierda de f^d coincide con el conjunto de discontinuidades por la izquierda de f.

En particular, se tiene:

$f^d(z) - f^d(x) = \sum_{\{y\in D:x<y\leq z\}} \left[f^d(y) - f^d(y-)\right]$ para cualquier parera (x, z) de números reales tales que $x < z$.

Así que f^d crece únicamente mediante saltos.

Por otra parte:

Si $x < z < 0$, se tiene:

$h(z) - h(x) = f(z) - f(x) - \left(f^d(z) - f^d(x)\right) = f(z) - f(x) - \sum_{\{y\in D:x<y\leq z\}} [f(y) - f(y-)]$.

Si $0 \leq x < z$, se tiene:

$h(z) - h(x) = f(z) - f(x) - \left(f^d(z) - f^d(x)\right) = f(z) - f(x) - \sum_{\{y\in D:x<y\leq z\}} [f(y) - f(y-)]$.

Si $x < 0$

$h(z) - h(x) = f(0) - f(x) - \left(f^d(0) - f^d(x)\right) = f(0) - f(x) - \sum_{\{y\in D:x<y\leq 0\}} [f(y) - f(y-)]$.

Así que, para cualquier parera (x, z) de números reales tales que $x < z$ se tiene:

$h(z) - h(x) = \sum_{\{y\in D:x<y\leq z\}} [f(y) - f(y-)] \geq 0$.

Por lo tanto, h es una función no decreciente.

$h(x) - h(x-) = f(x) - f(x-) - \left(f^d(x) - f^d(x-)\right) = 0.$

Así que h es continua por la izquierda.

Además:

$h(x+) - h(x) = f(x+) - f(x) - \left[f^d(x+) - f^d(x)\right] = f(x+) - f(x).$

Por el corolario 3.15 $h = f^i + f^c$, donde f^i es una función no decreciente, continua por la izquierda, que crece únicamente mediante saltos y tal que $f^i(x+) - f^i(x) = h(x+) - h(x)$ para cualquier $x \in \mathbb{R}$, y f^c es una función no decreciente y continua.

Se tiene entonces:

$f^i(x+) - f^i(x) = h(x+) - h(x) = f(x+) - f(x),$

$f = f^d + f^i + f^c.$

■

Corolario 3.13. *Toda función no decreciente $f : \mathbb{R} \to \mathbb{R}$ se puede expresar como la suma de dos funciones no decrecientes, f_1 y f_2; la primera, continua por la derecha y la segunda, continua por la izquierda y tales que, para cualquier $x \in \mathbb{R}$, $f_1(x) - f_1(x-) = f(x) - f(x-)$ y $f_2(x+) - f_2(x) = f(x+) - f(x)$.*

Corolario 3.14. *Sea $f : \mathbb{R} \to \mathbb{R}$ una función no decreciente y continua por la derecha, entonces f se puede expresar como la suma de una función f^d no decreciente, continua por la derecha, que crece únicamente mediante saltos y tal que $f^d(x) - f^d(x-) = f(x) - f(x-)$ para cualquier $x \in \mathbb{R}$, y una función f^c no decreciente y continua.*

Demostración

Sea $D = \{d_1, d_2, \ldots\}$ el conjunto de puntos donde f es discontinua y definamos $f^d : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:

$$f^d(x) = \begin{cases} f(0-) - \sum_{\{y \in D: x < y < 0\}} [f(y) - f(y-)] & \text{si } x \in (-\infty, 0) \\ f(0) + \sum_{\{y \in D: 0 < y \leq x\}} [f(y) - f(y-)] & \text{si } x \in [0, \infty) \end{cases}$$

Por el teorema 3.5, f^d es no decreciente, continua por la derecha, crece únicamente mediante saltos y $f^d(x) - f^d(x-) = f(x) - f(x-)$ para cualquier $x \in \mathbb{R}$.

Además, la función $f^c = f - f^d$ es no decreciente y continua por la izquierda.

Finalmente, como f y f^d son continuas por la derecha, f^c también lo es.

Por lo tanto, f^c es continua.

■

Corolario 3.15. *Sea $f : \mathbb{R} \to \mathbb{R}$ una función no decreciente y continua por la izquierda, entonces f se puede expresar como la suma de una función f^i no decreciente, continua por la izquierda, que crece únicamente mediante saltos y tal que $f^i(x+) - f^i(x) = f(x+) - f(x)$ para cualquier $x \in \mathbb{R}$, y una función f^c no decreciente y continua.*

Demostración

Definamos $h : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:

$h(x) = f(x+)$.

Por su definición, h es no decreciente y continua por la derecha. Además $h(x-) = f(x)$ para cualquier $x \in \mathbb{R}$.

Expresemos h como la suma de una función f^d no decreciente, continua por la derecha, que crece únicamente mediante saltos y tal que $f^d(x) - f^d(x-) = h(x) - h(x-)$ para cualquier $x \in \mathbb{R}$, y una función f^c no decreciente y continua.

Para cualquier $x \in \mathbb{R}$ se tiene:

$h(x) = f^d(x) + f^c(x)$.

Así que:

$f(x) = h(x-) = f^d(x-) + f^c(x)$.

Entonces, si $f^i : \mathbb{R} \to \mathbb{R}$ se define por $f^i(x) = f^d(x-)$, se tiene $f = f^i + f^c$.

Por su definición, f^i es no decreciente y continua por la izquierda.

Además, $f^i(x+) - f^i(x) = f^d(x) - f^d(x-)$ para cualquier $x \in \mathbb{R}$.

Como f^d crece únicamente mediante saltos, si (x, z) es una pareja de números reales tales que $x < z$, se tiene:

$f^d(z) - f^d(x) = \sum_{\{y \in D : x < y \leq z\}} \left[f^d(y) - f^d(y-)\right]$.

Así que:

$\sum_{\{y \in D : x \leq y < z\}} \left[f^i(y+) - f^i(y)\right] = \sum_{\{y \in D : x \leq y < z\}} \left[f^d(y) - f^d(y-)\right]$

$= \sum_{\{y \in D : x < y \leq z\}} \left[f^d(y) - f^d(y-)\right] - \left[f^d(z) - f^d(z-)\right] + \left[f^d(x) - f^d(x-)\right]$

$= \left[f^d(z) - f^d(x)\right] - \left[f^d(z) - f^d(z-)\right] + \left[f^d(x) - f^d(x-)\right]$

$= f^d(z-) - f^d(x-) = f^i(z) - f^i(x)$.

Por lo tanto, f^i crece únicamente mediante saltos.

Finalmente, para cualquier $x \in \mathbb{R}$, se tiene:

$f^i(x+) - f^i(x) = f^d(x) - f^d(x-) = h(x) - h(x-) = f(x+) - f(x)$.

■

Ahora extendemos los tres resultados anteriores a las funciones de variación acotada.

TEOREMA 3.6. *Toda función de variación acotada sobre cualquier intervalo compacto $g : \mathbb{R} \to \mathbb{R}$ se puede expresar como la suma de dos funciones de variación acotada sobre cualquier intervalo compacto, g_1 y g_2; la primera, continua por la derecha y la segunda, continua por la izquierda y tales que, para cualquier $x \in \mathbb{R}$,*

$g_1(x) - g_1(x-) = g(x) - g(x-)$ y $g_2(x+) - g_2(x) = g(x+) - g(x)$.

Demostración

Sean $f_1 : \mathbb{R} \to \mathbb{R}$ y $f_2 : \mathbb{R} \to \mathbb{R}$, no decrecientes y continuas por la derecha tales que $g = f_1 - f_2$.

Sean f_1^d, f_1^i, f_2^d y f_2^i funciones no decrecientes tales que $f_1 = f_1^d + f_1^i$, $f_2 = f_2^d + f_2^i$, f_1^d y f_2^d son continuas por la derecha, f_1^i y f_2^i son continuas por la izquierda, y tales que, para cualquier $x \in \mathbb{R}$, $f_1^d(x) - f_1^d(x-) = f_1(x) - f_1(x-)$, $f_1^i(x+) - f_1^i(x) = f_1(x+) - f_1(x)$, $f_2^d(x) - f_2^d(x-) = f_2(x) - f_2(x-)$, $f_2^i(x+) - f_2(x) = f_2(x+) - f_2(x)$.

Definamos $g_1 = f_1^d - f_2^d$ y $g_2 = f_1^i - f_2^i$

Entonces:

g_1 y g_2 son de variación acotada sobre cualquier intervalo compacto; g_1 es continua por la derecha y g_2 es continua por la izquierda.

Además:

$g = f_1 - f_2 = f_1^d + f_1^i - \left(f_2^d + f_2^i\right) = f_1^d - f_2^d + (f_1^i - f_2^i) = g_1 + g_2$

$g_1(x) - g_1(x-) = f_1^d(x) - f_2^d(x) - \left(f_1^d(x-) - f_2^d(x-)\right)$

$= f_1^d(x) - f_1^d(x-) - \left(f_2^d(x) - f_2^d(x-)\right) = f_1(x) - f_1(x-) - (f_2(x) - f_2(x-))$

$= g(x) - g(x-)$

$g_2(x+) - g_2(x) = f_1^i(x+) - f_2^i(x+) - (f_1^i(x) - f_2^i(x))$

$= f_1^i(x+) - f_1^i(x) - (f_2^i(x+) - f_2^i(x)) = f_1(x+) - f_1(x) - (f_2(x+) - f_2(x))$

$= g(x+) - g(x)$

■

Teorema 3.7. *Sea $g : \mathbb{R} \to \mathbb{R}$ una función continua por la derecha y de variación acotada sobre cualquier intervalo compacto, entonces g se puede expresar como la suma de dos funciones de variación acotada sobre cualquier intervalo compacto, g^d y g^c; la primera, continua por la derecha, que crece o decrece únicamente mediante saltos y tal que:*

$g^d(x) - g^d(x-) = g(x) - g(x-)$

para cualquier $x \in \mathbb{R}$, y la segunda, continua.

Demostración

Sean $f_1 : \mathbb{R} \to \mathbb{R}$ y $f_2 : \mathbb{R} \to \mathbb{R}$, no decrecientes y continuas por la derecha tales que $g = f_1 - f_2$.

Por el teorema 3.14, se tiene:

$f_1 = f_1^d + f_1^c$,

$f_1^d(x) - f_1^d(x-) = f_1(x) - f_1(x-)$ para cualquier $x \in \mathbb{R}$,

$f_1^d(z) - f_1^d(x) = \sum_{\{y \in D : x < y \leq z\}} \left[f_1^d(y) - f_1^d(y-)\right]$ para cualquier pareja (x, z) de números reales tales que $x < z$,

$f_2 = f_2^d + f_2^c$,

$f_2^d(x) - f_2^d(x-) = f_2(x) - f_2(x-)$, para cualquier $x \in \mathbb{R}$.

$f_2^d(z) - f_2^d(x) = \sum_{\{y \in D : x < y \leq z\}} \left[f_2^d(y) - f_2^d(y-)\right]$, para cualquier parera (x, z) de números reales tales que $x < z$,

donde f_1^c y f_2^c son funciones no decrecientes continuas y f_1^d y f_2^d son funciones no decrecientes, continuas por la derechas y que crecen únicamente mediante saltos.

Así que:

$g = \left(f_1^d + f_1^c\right) - \left(f_2^d + f_2^c\right) = \left(f_1^d - f_2^d\right) + \left(f_1^c - f_2^c\right).$

Definamos:

$g^d = f_1^d - f_2^d$,

$g^c = f_1^c - f_2^c$.

g^d es continua por la derecha y de variación acotada sobre cualquier intervalo compacto.

g^c es continua y de variación acotada sobre cualquier intervalo compacto.

Para cualquier $x \in \mathbb{R}$, se tiene:

$g^d(x) - g^d(x-) = f_1^d(x) - f_1^d(x-) - \left[f_2^d(x) - f_2^d(x-)\right]$

$= f_1(x) - f_1(x-) - [f_2(x) - f_2(x-)] = f_1(x) - f_2(x) - [f_1(x-) - f_2(x-)]$

$= g(x) - g(x-)$.

Como $g^d = f_1^d - f_2^d$, g^d puede crecer en un intervalo $[x,z]$, donde $x < z$, únicamente si f_1^d crece en el intervalo $[x,z]$. Además, f_1^d crece únicamente mediante saltos. Si f_1^d crece mediante un salto $f_1^d(y) - f_1^d(y-)$, entonces g^d crece mediante la diferencia de los saltos $f_1^d(y) - f_1^d(y-)$ y $f_2^d(y) - f_2^d(y-)$, cuando esta diferencia es positiva. Además, se tiene:

$$f_1^d(y) - f_1^d(y-) - \left[f_2^d(y) - f_2^d(y-)\right] = f_1^d(y) - f_2^d(y) - \left[f_1^d(y-) - f_2^d(y-)\right] = g^d(y) - g^d(y-).$$

Por lo tanto, g^d crece en el intervalo $[x,z]$ únicamente mediante los saltos $g^d(y) - g^d(y-)$ que sean positivos, donde $y \in (x,z]$.

De la misma manera, g^d decrece en el intervalo $[x,z]$ únicamente mediante los saltos $g^d(y) - g^d(y-)$ que sean negativos, donde $y \in (x,z]$.

Para cualquier parera (x,z) de números reales tales que $x < z$, se tiene:

$g^d(z) - g^d(x) = f_1^d(z) - f_1^d(x) - \left[f_2^d(z) - f_2^d(x)\right]$

$= \sum_{\{y \in D: x < y \leq z\}} \left[f_1^d(y) - f_1^d(y-)\right] - \sum_{\{y \in D: x < y \leq z\}} \left[f_2^d(y) - f_2^d(y-)\right]$

$= \sum_{\{y \in D: x < y \leq z\}} \left[\left(f_1^d(y) - f_2^d(y)\right) - \left(f_1^d(y-) - f_2^d(y-)\right)\right]$

$= \sum_{\{y \in D: x < y \leq z\}} \left[g^d(y) - g^d(y-)\right]$.

■

Corolario 3.16. *Sea $g : \mathbb{R} \to \mathbb{R}$ una función continua por la izquierda y de variación acotada sobre cualquier intervalo compacto, entonces g se puede expresar como la suma de dos funciones de variación acotada sobre cualquier intervalo compacto, g^i y g^c; la primera, continua por la izquierda, que crece o decrece únicamente mediante saltos y tal que:*

$g^i(x+) - g^i(x) = g(x+) - g(x)$

para cualquier $x \in \mathbb{R}$, y la segunda, continua.

A continuación exponemos el resultado general de descomposición de una función no decreciente en su parte continua y su parte de saltos. Ésta última puede descoponerse en una parte de saltos por la derecha más una parte de saltos por la izquierda.

Teorema 3.8. *Sea $f : \mathbb{R} \to \mathbb{R}$ una función no decreciente, entonces f se puede expresar como la suma de tres funciones no decrecientes, f^d, f^i y f^c; la primera, continua por la derecha, que crece únicamente mediante saltos y tal que $f^d(x) - f^d(x-) = f(x) - f(x-)$ para cualquier $x \in \mathbb{R}$, la segunda, continua por la izquierda, que crece únicamente mediante saltos y tal que $f^i(x+) - f^i(x) = f(x+) - f(x)$ para cualquier $x \in \mathbb{R}$, y la tercera, continua.*

Demostración

Sea $D = \{d_1, d_2, \ldots\}$ el conjunto de puntos donde f es discontinua y definamos $f^d : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:

$$f^d(x) = \begin{cases} f(0-) - \sum_{\{y \in D : x < y < 0\}} [f(y) - f(y-)] & \text{si } x \in (-\infty, 0) \\ f(0) + \sum_{\{y \in D : 0 < y \leq x\}} [f(y) - f(y-)] & \text{si } x \in [0, \infty) \end{cases}$$

Por el teorema 3.5, se tiene:

f^d es una función no decreciente, continua por la derecha, crece únicamente mediante saltos y $f^d(x) - f^d(x-) = f(x) - f(x-)$ para cualquier $x \in \mathbb{R}$.

Además, la función $h = f - f^d$ es no decreciente, continua por la izquierda y $h(x+) - h(x) = f(x+) - f(x)$ para cualquier $x \in \mathbb{R}$.

Por el corolario 3.15 $h = f^i + f^c$, donde f^i es una función no decreciente, continua por la izquierda, que crece únicamente mediante saltos y tal que $f^i(x+) - f^i(x) = h(x+) - h(x)$ para cualquier $x \in \mathbb{R}$, y f^c es una función no decreciente y continua.

Se tiene entonces:

$f^i(x+) - f^i(x) = h(x+) - h(x) = f(x+) - f(x)$,

$f = f^d + h = f^d + f^i + f^c$.

■

CAPÍTULO 4

LA INTEGRAL DE STIELTJES

4.1. La integral de Riemann-Stieltjes

Así como se define la integral de Riemann, $\int_a^b f(x)dx$, de una función acotada $f : [a, b] \to \mathbb{R}$, puede definirse la integral de una función f con respecto a una función g. Como lo mencionamos en la introducciuón de este capítulo, este problema fue abordado y resuelto por Stieltjes a finales del siglo XIX.

Para la definición general no es necesario restringirse al caso de una función g de variación acotada. La integral de Stieltjes puede estar bien definida aún para el caso de una función g que no es de variación acotada. La necesidad de integrar con respecto a una función g de variación acotada surge si queremos que todas las funciones continuas sean integrables con respecto a g, propiedad que únicamente se tiene cuando g es de variación acotada.

Por otra parte, si queremos generar una medida a partir de una función g, también es necesario que ésta sea de variación acotada.

La definición y las propiedades de la integral de Riemann-Stieltjes son similares a las de la integral de Riemann. Ésta última resulta más simple ya que la función con respecto a la cual se integra es creciente y continua.

DEFINICIÓN 4.1. *Sean $f : [a, b] \to \mathbb{R}$ y $g : [a, b] \to \mathbb{R}$ dos funciones acotadas y $P = \{x_0, x_1, \ldots, x_n\}$ una particion del intervalo $[a, b]$. Una suma de Riemann-Stieltjes $S(P, f, g)$ de f con respecto a g, correspondiente a la particion P, es una suma de la forma $S(P, f, g) = \sum_{k=1}^{n} f(\xi_k)\,[g(x_k) - g(x_{k-1})]$, donde $\xi_k \in [x_{k-1}, x_k]$ para $k \in \{1, 2, \ldots, n\}$.*

DEFINICIÓN 4.2. *Se dice que f es integrable con respecto a g en el intervalo $[a, b]$ si existe un numero real I tal que para cualquier $\varepsilon > 0$ existe una particion P_ε del intervalo $[a, b]$ tal que $|S(P, f, g) - I| < \varepsilon$ para cualquier particion P que sea un refinamiento de P_ε y cualquier suma de Riemann-Stieltjes $S(P, f, g)$ de f con respecto a g, correspondiente a la particion P. Al numero real I de esta definición se le llama la integral de f con respecto a g y se le denota por $\int_a^b f dg$.*

4.2. Criterio de Cauchy

DEFINICIÓN 4.3. *Se dice que la pareja de funciones acotadas $f : [a, b] \to \mathbb{R}$ y $g : [a, b] \to \mathbb{R}$ satisface el criterio de Cauchy si para cada $\varepsilon > 0$ existe una partición P_ε del intervalo $[a, b]$ tal que si P y P' son dos refinamientos de P_ε y $S(P, f, g)$, $S(P', f, g)$ son sumas de Riemann-Stieltjes de f con respecto a g, entonces:*

$|S(P, f, g) - S(P', f, g)| < \varepsilon$

TEOREMA 4.1. *Una función f es integrable con respecto a g en el intervalo $[a, b]$ si y sólo si la pareja f, g satisface el criterio de Cauchy.*

Demostración

Si f es integrable con respecto a g, claramente se satisface el criterio de Cauchy. Para el inverso, definamos inductivamente una sucesión de particiones $\{Q_n\}$ tal que $Q_0 = \{a, b\}$ y, para $n \geq 1$, $Q_n \supset Q_{n-1}$ y si $P, Q \supset Q_n$ entonces $|S(P, f, g) - S(Q, f, g)| < \frac{1}{n}$ para cualquier par de sumas de Riemann-Stieltjes $S(P, f, g)$ y $S(Q, f, g)$.

Para cada n consideremos entonces cualquier suma de Riemann-Stieltjes $S(Q_n, f, g)$. La sucesión $\{S(Q_n, f, g) : n \in \mathbb{N}\}$ claramente es de Cauchy y por lo tanto converge.

Sea $I = \text{lím}_{n \rightsquigarrow \infty} S(Q_n, f, g)$ y, dada $\varepsilon > 0$, sea N tal que $\frac{1}{N} < \frac{\varepsilon}{2}$ y $|S(Q_n, f, g) - I| < \frac{\varepsilon}{2}$ para toda $n \geq N$.

Si $P \supset Q_N$, se tiene $|S(P, f, g) - S(Q_N, f, g)| < \frac{1}{N}$, por lo tanto:

$|S(P, f, g) - I| \leq |S(P, f, g) - S(Q_N, f, g)| + |S(Q_N, f, g) - I| < \frac{1}{N} + \frac{\varepsilon}{2} < \varepsilon.$

Se concluye entonces que f es integrable con respecto a g.

■

PROPOSICIÓN 4.1. *Si f es integrable con respecto a g, entonces f y g no tienen discontinuidades en común del mismo lado.*

Demostración

Supongamos que f no es continua por la izquierda en $c \in (a, b]$. Existe entonces $M > 0$ tal que para cualquier $\delta > 0$ existe $y \in (c - \delta, c)$ tal que $|f(c) - f(y)| > M$.

Dada $\varepsilon > 0$, sea P_ε una partición del intervalo $[a, b]$ tal que $|S(P, f, g) - S(P', f, g)| < M\varepsilon$ para todo par de refinamientos P y P' de P_ε. Si el punto c no forma parte de la partición P_ε se le puede agregar y la nueva partición sigue teniendo la misma propiedad que P_ε, de manera que podemos siempre elegir $P\varepsilon$ de tal manera que contenga al punto c.

Sea d el punto de $P\varepsilon$ que se encuentre inmediatamente a la izquierda de c. Entonces, si $x \in (d, c)$ definamos las particiones $P = Q = P_\varepsilon \cup \{x\}$ y consideremos dos sumas de Riemann-Stieltjes $S(P, f, g)$ y $S(Q, f, g)$ de la siguiente manera:

En el subintervalo $[x, c]$, al definir la suma $S(P, f, g)$ elijamos el punto c como punto intermedio, mientras que al definir la suma $S(Q, f, g)$ elijamos un punto ξ tal que $|f(c) - f(\xi)| > M$. En los otros subintervalos de la partición elijamos el mismo punto intermedio para definir cualquiera de las dos sumas de Riemann-Stieltjes $S(P, f, g)$ y $S(Q, f, g)$. Se tiene entonces:

$|[f(c) - f(\xi)]\,[g(c) - g(x)]| = |S(P, f, g) - S(Q, f, g)| < M\varepsilon.$

Por lo tanto, $|g(c) - g(x)| < \varepsilon$.

Se puede concluir entonces que g es continua por la izquierda en c.

De la misma manera, se tiene que si f no es continua por la derecha en $c \in [a, b)$, entonces g sí lo es.

■

Teorema 4.2. *Si g es de variación acotada, entonces toda función continua es integrable con respecto a g.*

Demostración

Sea $g : [a, b] \to \mathbb{R}$ de variación acotada y $f : [a, b] \to \mathbb{R}$ continua.

Si g es constante en el intervalo $[a, b]$, el resultado es trivial.

Supongamos que g no es constante en el intervalo $[a, b]$ y definamos $v = V_g[a, b]$.

Como f es uniformemente continua en $[a, b]$, dada $\varepsilon > 0$ existe $\delta > 0$ tal que si $x, y \in [a, b]$ y $|x - y| < \delta$, entonces $|f(x) - f(y)| < \frac{\varepsilon}{2v}$.

Sea P_ε una partición de $[a, b]$ de norma menor que δ, P un refinamiento de P_ε y $S(P, f, g)$, $S(P_\varepsilon, f, g)$ sumas de Riemann-Stieltjes de f con respecto a g.

Si $P_\varepsilon = \{x_0, x_1, \ldots, x_n\}$, $P = \{y_0, y_1, \ldots, y_m\}$ y $S(P, f, g) = \sum_{j=1}^{m} f(\xi_j)\,[g(y_j) - g(y_{j-1})]$, entonces, por ser P un refinamiento de P_ε, $S(P_\varepsilon, f, g)$ puede escribirse en la forma:

$S(P_\varepsilon, f, g) = \sum_{j=1}^{m} f(\beta_j)\,[g(y_j) - g(y_{j-1})]$,

donde β_j no necesariamente pertenece al intervalo $[y_{j-1}, y_j]$, pero de tal manera que, para cada $j \in \{1, 2, \ldots, m\}$, ξ_j y β_j pertenecen a un mismo intervalo de la forma $[x_{k-1}, x_k]$. Se tiene entonces:

$|S(P, f, g) - S(P_\varepsilon, f, g)| = \left|\sum_{j=1}^{m} \left[f(\xi_j) - f(\beta_j)\right] [g(y_j) - g(y_{j-1})]\right|$

$< \frac{\varepsilon}{2v} \left|\sum_{j=1}^{m} [g(y_j) - g(y_{j-1})]\right| \leq \frac{\varepsilon}{2v} V_g[a, b] = \frac{\varepsilon}{2}.$

Si P' es otro refinamiento de P_ε y $S(P', f, g)$ es una suma de Riemann-Stieltjes de f con respecto a g, se tiene:

$|S(P,f,g) - S(P',f,g)| \leq |S(P,f,g) - S(P_\varepsilon,f,g)| + |S(P',f,g) - S(P_\varepsilon,f,g)| < \varepsilon.$

Así que la pareja f, g satisface el criterio de Cauchy y, por lo tanto, f es integrable con respecto a g.

∎

TEOREMA 4.3. *Si* $f_1 : [a,b] \to \mathbb{R}$ *y* $f_2 : [a,b] \to \mathbb{R}$ *son integrables con respecto a* $g : [a,b] \to \mathbb{R}$ *y* $\alpha_1, \alpha_2 \in \mathbb{R}$*, entonces* $\alpha_1 f_1 + \alpha_2 f_2$ *es integrable con respecto a* g *y:*

$\int_a^b (\alpha_1 f_1 + \alpha_2 f_2)\, dg = \alpha_1 \int_a^b f_1 dg + \alpha_2 \int_a^b f_2 dg.$

Demostración

Si $f : [a,b] \to \mathbb{R}$ es integrable con respecto a g, se prueba inmediatamente que si $\alpha \in \mathbb{R}$, entonces αf es integrable con respecto a g y que $\int_a^b \alpha f dg = \alpha \int_a^b f dg$. Así que, para demostrar la proposición, basta con demostrar que $f_1 + f_2$ es integrable con respecto a g y que $\int_a^b (f_1 + f_2)\, dg = \int_a^b f_1 dg + \int_a^b f_2 dg$.

Dada $\varepsilon > 0$ sea $P_\varepsilon^{(1)}$ (resp. $P_\varepsilon^{(2)}$) una partición del intervalo $[a,b]$ tal que:

$\left| S(P,f_1,g) - \int_a^b f_1 dg \right| < \frac{1}{2}\varepsilon$ (resp. $\left| S(P,f_2,g) - \int_a^b f_2 dg \right| < \frac{1}{2}\varepsilon$)

para cualquier particion P que sea un refinamiento de $P_\varepsilon^{(1)}$ (resp. $P_\varepsilon^{(2)}$) y cualquier suma de Riemann-Stieltjes $S(P,f_1,g)$ (resp. $S(P,f_2,g)$) de f_1 (resp. f_2) con respecto a g, correspondiente a la partición P.

Definamos $P_\varepsilon = P_\varepsilon^{(1)} \cup P_\varepsilon^{(2)}$ y sean $P = \{x_0, x_1, \ldots, x_n\}$ un refinamiento de P_ε y:

$S(P, f_1 + f_2, g) = \sum_{k=1}^n [f_1(\xi_k) + f_2(\xi_k)]\, [g(x_k) - g(x_{k-1})]$

una suma de Riemann-Stieltjes de $f_1 + f_2$ con respecto a g, correspondiente a la partición P. Entonces P es un refinamiento de $P_\varepsilon^{(1)}$ y de $P_\varepsilon^{(2)}$,

$S(P,f_1,g) = \sum_{k=1}^n f_1(\xi_k)\, [g(x_k) - g(x_{k-1})]$ (resp. $S(P,f_2,g) = \sum_{k=1}^n f_2(\xi_k)\, [g(x_k) - g(x_{k-1})]$

es una suma de Riemann-Stieltjes de f_1 (resp. f_2) con respecto a g, correspondiente a la partición P, y se tiene:

$\left| S(P, f_1 + f_2, g) - \int_a^b f_1 dg - \int_a^b f_2 dg \right| = \left| S(P,f_1,g) + S(P,f_2,g) - \int_a^b f_1 dg - \int_a^b f_2 dg \right|$

$\leq \left| S(P,f_1,g) - \int_a^b f_1 dg \right| + \left| S(P,f_2,g) - \int_a^b f_2 dg \right| < \varepsilon.$

Por lo tanto, $f_1 + f_2$ es integrable con respecto a g y $\int_a^b (f_1 + f_2)\, dg = \int_a^b f_1 dg + \int_a^b f_2 dg$.

∎

Teorema 4.4. *Si $f : [a,b] \to \mathbb{R}$ es integrable con respecto a $g_1 : [a,b] \to \mathbb{R}$ y con respecto a $g_2 : [a,b] \to \mathbb{R}$ y $\alpha_1, \alpha_2 \in \mathbb{R}$, entonces f es integrable con respecto a $\alpha_1 g_1 + \alpha_2 g_2$ y:*

$\int_a^b f d\left(\alpha_1 g_1 + \alpha_2 g_2\right) = \alpha_1 \int_a^b f dg_1 + \alpha_2 \int_a^b f dg_2.$

Demostración

Si $f : [a,b] \to \mathbb{R}$ es integrable con respecto a g, se prueba inmediatamente que si $\alpha \in \mathbb{R}$, entonces f es integrable con respecto a αg y que $\int_a^b f d\left(\alpha g\right) = \alpha \int_a^b f dg$. Así que, para demostrar la proposición, basta con demostrar que f es integrable con respecto a $g_1 + g_2$ y que $\int_a^b f d\left(g_1 + g_2\right) = \int_a^b f dg_1 + \int_a^b f dg_2$.

Dada $\varepsilon > 0$ sea $P_\varepsilon^{(1)}$ (resp. $P_\varepsilon^{(2)}$) una partición del intervalo $[a,b]$ tal que:

$\left|S(P,f,g_1) - \int_a^b f dg_1\right| < \frac{1}{2}\varepsilon$ (resp. $\left|S(P,f,g_2) - \int_a^b f dg_2\right| < \frac{1}{2}\varepsilon$)

para cualquier partición P que sea un refinamiento de $P_\varepsilon^{(1)}$ (resp. $P_\varepsilon^{(2)}$) y cualquier suma de Riemann-Stieltjes $S(P,f,g_1)$ (resp. $S(P,f,g_2)$) de f con respecto a g_1 (resp. g_2), correspondiente a la partición P.

Definamos $P_\varepsilon = P_\varepsilon^{(1)} \cup P_\varepsilon^{(2)}$ y sean $P = \{x_0, x_1, \ldots, x_n\}$ un refinamiento de P_ε y:

$S(P,f,g_1+g_2) = \sum_{k=1}^n f(\xi_k)\left[\left(g_1+g_2\right)(x_k) - \left(g_1+g_2\right)(x_{k-1})\right]$

una suma de Riemann-Stieltjes de f con respecto a $g_1 + g_2$, correspondiente a la partición P. Entonces P es un refinamiento de $P_\varepsilon^{(1)}$ y de $P_\varepsilon^{(2)}$,

$S(P,f,g_1) = \sum_{k=1}^n f(\xi_k)\left[g_1(x_k) - g_1(x_{k-1})\right]$ (resp. $S(P,f,g_2) = \sum_{k=1}^n f(\xi_k)\left[g_2(x_k) - g_2(x_{k-1})\right]$)

es una suma de Riemann-Stieltjes de f con respecto a g_1 (resp. g_2), correspondiente a la partición P, y se tiene:

$\left|S(P,f,g_1+g_2) - \int_a^b f dg_1 - \int_a^b f dg_2\right| = \left|S(P,f,g_1) + S(P,f,g_2) - \int_a^b f dg_1 - \int_a^b f dg_2\right|$

$\leq \left|S(P,f,g_1) - \int_a^b f dg_1\right| + \left|S(P,f,g_2) - \int_a^b f dg_2\right| < \varepsilon.$

Por lo tanto, f es integrable con respecto a $g_1 + g_2$ y $\int_a^b f d\left(g_1+g_2\right) = \int_a^b f dg_1 + \int_a^b f dg_2$.

■

Teorema 4.5. *Sean $f : [a,b] \to \mathbb{R}$ y $g : [a,b] \to \mathbb{R}$ funciones acotadas y $c \in [a,b]$, entonces f es integrable con respecto a g en el intervalo $[a,b]$ si y sólo si es integrable con respecto a g en cada uno de los intervalos $[a,c]$ y $[c,b]$. En ese caso se tiene:*

$\int_a^b f dg = \int_a^c f dg + \int_c^b f dg.$

Demostración

Si $c \in \{a, b\}$, el resultado es inmediato, así que resta únicamente probar la proposición cuando $c \in (a, b)$.

Supongamos primero que f es integrable con respecto a g en el intervalo $[a, b]$.

Dada $\varepsilon > 0$ sea $P_\varepsilon = \{x_0, x_1, \dots, x_n\}$ una partición del intervalo $[a, b]$ tal que:

$|S(P, f, g) - S(P', f, g)| < \varepsilon$

para cualquier par de particiones P y P' que sean refinamientos de P_ε y cualquier suma de Riemann-Stieltjes $S(P, f, g)$ (resp. $S(P', f, g)$) de f con respecto a g, correspondiente a la partición P (resp. P').

Definamos $P_\varepsilon^{(1)} = \{x_0, x_1, \dots, x_j, c\}$ y $P_\varepsilon^{(2)} = \{c, x_{j+1}, \dots, x_n\}$.

Sean $P_1^{(1)}$ y $P_2^{(1)}$ dos refinamientos de $P_\varepsilon^{(1)}$, $P^{(2)}$ un refinamiento de $P_\varepsilon^{(2)}$, $S(P_1^{(1)}, f, g)$ una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la particion $P_1^{(1)}$, $S(P_2^{(1)}, f, g)$ una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la particion $P_2^{(1)}$ y $S(P^{(2)}, f, g)$ una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la particion $P^{(2)}$. Entonces $P_1^{(1)} \cup P^{(2)}$ y $P_2^{(1)} \cup P^{(2)}$ son refinamientos de P_ε, $S(P_1^{(1)}, f, g) + S(P^{(2)}, f, g)$ es una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la particion $P_1^{(1)} \cup P^{(2)}$ y $S(P_2^{(1)}, f, g) + S(P^{(2)}, f, g)$ es una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la particion $P_2^{(1)} \cup P^{(2)}$. Por lo tanto:

$$\left| S(P_1^{(1)}, f, g) - S(P_2^{(1)}, f, g) \right|$$

$$= \left| S(P_1^{(1)}, f, g) + S(P^{(2)}, f, g) - S(P_2^{(1)}, f, g) - S(P^{(2)}, f, g) \right| < \varepsilon.$$

Así que, por el criterio de Cauchy, f es integrable con respecto a g en el intervalo $[a, c]$.

De manera análoga se demuestra que f es integrable con respecto a g en el intervalo $[c, b]$.

Inversamente, supongamos que f es integrable con respecto a g en cada uno de los intervalos $[a, c]$ y $[c, b]$.

Dada $\varepsilon > 0$ sea $P_\varepsilon^{(1)} = \{x_0, x_1, \dots, c\}$ una partición del intervalo $[a, c]$ tal que:

$$\left| S(P_1^{(1)}, f, g) - S(P_2^{(1)}, f, g) \right| < \tfrac{1}{2}\varepsilon$$

para cualquier par de particiones $P_1^{(1)}$ y $P_2^{(1)}$ que sean refinamientos de $P_\varepsilon^{(1)}$ y cualquier suma de Riemann-Stieltjes $S(P_1^{(1)}, f, g)$ (resp. $S(P_2^{(1)}, f, g)$) de f con respecto a g, correspondiente a la partición $P_1^{(1)}$ (resp. $P_2^{(1)}$).

Sea también $P_\varepsilon^{(2)} = \{c, y_1, \ldots, y_n\}$ una partición del intervalo $[c, b]$ tal que:

$$\left|S(P_1^{(2)}, f, g) - S(P_2^{(2)}, f, g)\right| < \tfrac{1}{2}\varepsilon$$

para cualquier par de particiones $P_1^{(2)}$ y $P_2^{(2)}$ que sean refinamientos de $P_\varepsilon^{(2)}$ y cualquier suma de Riemann-Stieltjes $S(P_1^{(2)}, f, g)$ (resp. $S(P_2^{(2)}, f, g)$) de f con respecto a g, correspondiente a la partición $P_1^{(2)}$ (resp. $P_2^{(2)}$).

Definamos $P_\varepsilon = P_\varepsilon^{(1)} \cup P_\varepsilon^{(2)}$ y sean P y P' dos refinamientos de P_ε y $S(P, f, g)$ (resp. $S(P', f, g)$) una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la particion P (resp. P'). Entonces $P_1^{(1)} = P \cap [a, c]$ y $P_2^{(1)} = P' \cap [a, c]$ son refinamientos de $P_\varepsilon^{(1)}$ y $P_1^{(2)} = P \cap [c, b]$ y $P_2^{(2)} = P' \cap [c, b]$ son refinamientos de $P_\varepsilon^{(2)}$. Además, $S(P, f, g)$ y $S(P', f, g)$) se pueden expresar de la siguiente manera:

$$S(P, f, g) = S(P_1^{(1)}, f, g) + S(P_1^{(2)}, f, g),$$

$$S(P', f, g) = S(P_2^{(1)}, f, g) + S(P_2^{(2)}, f, g),$$

donde $S(P_1^{(1)}, f, g)$ (resp. $S(P_1^{(2)}, f, g)$) es una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la partición $P_1^{(1)}$ (resp. $P_1^{(2)}$), y $S(P_2^{(1)}, f, g)$ (resp. $S(P_2^{(2)}, f, g)$) es una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la partición $P_2^{(1)}$ (resp. $P_2^{(2)}$).

Por lo tanto:

$$|S(P, f, g) - S(P', f, g)| = \left|S(P_1^{(1)}, f, g) + S(P_1^{(2)}, f, g) - S(P_2^{(1)}, f, g) - S(P_2^{(2)}, f, g)\right|$$

$$\leq \left|S(P_1^{(1)}, f, g) - S(P_2^{(1)}, f, g)\right| + \left|S(P_1^{(2)}, f, g) - S(P_2^{(2)}, f, g)\right| < \varepsilon.$$

Así que, por el criterio de Cauchy, f es integrable con respecto a g en el intervalo $[a, b]$.

Finalmente, si f es integrable con respecto a g en cada uno de los intervalos $[a, c]$ y $[c, b]$, dada $\varepsilon > 0$, sea $P_\varepsilon^{(1)}$ (resp. $P_\varepsilon^{(2)}$) una partición del intervalo $[a, c]$ (resp. $[c, b]$) tal que $\left|S\left(P^{(1)}, f, g\right) - \int_a^c f dg\right| < \frac{1}{2}\varepsilon$ (resp. $\left|S(P^{(2)}, f, g) - \int_c^b f dg\right| < \frac{1}{2}\varepsilon$) para cualquier partición $P^{(1)}$ (resp. $P^{(2)}$) que sea un refinamiento de $P_\varepsilon^{(1)}$ (resp. $P_\varepsilon^{(2)}$) y cualquier suma de Riemann-Stieltjes $S\left(P^{(1)}, f, g\right)$ (resp. $S(P^{(2)}, f, g)$) de f con respecto a g correspondiente a la partición $P^{(1)}$ (resp. $P^{(2)}$).

Definamos $P_\varepsilon = P_\varepsilon^{(1)} \cup P_\varepsilon^{(2)}$ y sean P un refinamiento de P_ε y $S(P, f, g)$ una suma de Riemann-Stieltjes de f con respecto a g correspondiente a la partición P. Entonces $P^{(1)} = P \cap [a, c]$ (resp. $P^{(2)} = P \cap [c, b]$) es un refinamiento de $P_\varepsilon^{(1)}$ (resp. $P_\varepsilon^{(2)}$) y $S(P, f, g)$ se puede expresar de la siguiente manera:

$S(P, f, g) = S(P^{(1)}, f, g) + S(P^{(2)}, f, g)$,

donde $S(P^{(1)}, f, g)$ (resp. $S(P^{(2)}, f, g)$) es una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la partición $P^{(1)}$ (resp. $P^{(2)}$).

Por lo tanto:

$$\left|S(P, f, g) - \int_a^c f dg - \int_c^b f dg\right| = \left|S(P^{(1)}, f, g) + S(P^{(2)}, f, g) - \int_a^c f dg - \int_c^b f dg\right|$$

$$\leq \left|S(P^{(1)}, f, g) - \int_a^c f dg\right| + \left|S(P^{(2)}, f, g) - \int_c^b f dg\right| < \varepsilon.$$

Así que:

$$\int_a^b f dg = \int_a^c f dg + \int_c^b f dg$$

■

Proposición 4.2. *Si $f : [a, b] \to \mathbb{R}$ es una función no negativa e integrable con respecto a una función no decreciente $g : [a, b] \to \mathbb{R}$, entonces $\int_a^b f dg \geq 0$.*

Demostración

Supongamos que $\int_a^b f dg < 0$ y sea $\varepsilon \in \left(0, \left|\int_a^b f dg\right|\right)$.

Sea P_ε una partición del intervalo $[a, b]$ tal que $\left|S(P, f, g) - \int_a^b f dg\right| < \varepsilon$ para cualquier particion P que sea un refinamiento de P_ε y cualquier suma de Riemann-Stieltjes $S(P, f, g)$ de f con respecto a g, correspondiente a la particion P.

Como f es no negativa y g es no decreciente, $S(P, f, g) \geq 0$, así que:

$$S(P, f, g) - \int_a^b f dg = \left|S(P, f, g) - \int_a^b f dg\right| < \varepsilon < \left|\int_a^b f dg\right| = -\int_a^b f dg.$$

Por lo tanto, $S(P, f, g) < 0$, lo cual es una contradicción.

■

Corolario 4.1. *Si $f_1 : [a, b] \to \mathbb{R}$ y $f_2 : [a, b] \to \mathbb{R}$ son dos funciones integrables con respecto a una función no decreciente $g : [a, b] \to \mathbb{R}$ y $f_1 \leq f_2$, entonces $\int_a^b f_1 dg \leq \int_a^b f_2 dg$.*

Proposición 4.3. *Si $f : [a, b] \to \mathbb{R}$ es integrable con respecto a una función no decreciente $g : [a, b] \to \mathbb{R}$, entonces $|f|$ es integrable con respecto a g y:*

$$\left|\int_a^b f dg\right| \leq \int_a^b |f|\, dg.$$

Demostración

Dada $\varepsilon > 0$, sea P_ε una partición del intervalo $[a, b]$ tal que si P y P' son dos refinamientos de P_ε y $S(P, f, g)$, $S(P', f, g)$ son sumas de Riemann-Stieltjes de f con respecto a g, entonces $|S(P, f, g) - S(P', f, g)| < \frac{1}{2}\varepsilon$.

Sean P un refinamiento de P_ε y:

$S(P, |f|, g) = \sum_{k=1}^{n} |f(\xi_k)| [g(x_k) - g(x_{k-1})]$,

$S(P_\varepsilon, |f|, g) = \sum_{k=1}^{m} |f(\eta_k)| [g(z_k) - g(z_{k-1})]$,

dos sumas de Riemann-Stieltjes de $|f|$ con respecto a g.

Como P es un refinamiento de P_ε, cada intervalo $[x_{i-1}, x_i]$ está contenido en un intervalo $[z_{j-1}, z_j]$, definamos entonces $\beta_i = \eta_j$ si $[x_{i-1}, x_i] \subset [z_{j-1}, z_j]$. Entonces:

$S(P_\varepsilon, |f|, g) = \sum_{k=1}^{n} |f(\beta_k)| [g(x_k) - g(x_{k-1})]$.

Si $f(\xi_k) - f(\beta_k) \geq 0$, tomemos $\alpha_k = \xi_k$ y $\gamma_k = \beta_k$.

Si $f(\xi_k) - f(\beta_k) < 0$, tomemos $\alpha_k = \beta_k$ y $\gamma_k = \xi_k$.

Tomemos:

$S(P, f, g) = \sum_{k=1}^{n} f(\alpha_k) [g(x_k) - g(x_{k-1})]$,

$S'(P, f, g) = \sum_{k=1}^{n} f(\gamma_k) [g(x_k) - g(x_{k-1})]$.

Entonces se tiene:

$\left|S(P, f, g) - S'(P, f, g)\right| = \left|\sum_{k=1}^{n} [f(\alpha_k) - f(\gamma_k)] [g(x_k) - g(x_{k-1})]\right|$

$= \sum_{k=1}^{n} [f(\alpha_k) - f(\gamma_k)] [g(x_k) - g(x_{k-1})] = \sum_{k=1}^{n} |f(\xi_k) - f(\beta_k)| [g(x_k) - g(x_{k-1})]$.

Así que:

$\sum_{k=1}^{n} |f(\xi_k) - f(\beta_k)| [g(x_k) - g(x_{k-1})] < \frac{1}{2}\varepsilon$.

Por lo tanto:

$|S(P, |f|, g) - S(P_\varepsilon, |f|, g)| = \left|\sum_{k=1}^{n} (|f(\xi_k)| - |f(\beta_k)|) [g(x_k) - g(x_{k-1})]\right|$

$\leq \sum_{k=1}^{n} |f(\xi_k) - f(\beta_k)| [g(x_k) - g(x_{k-1})] < \frac{1}{2}\varepsilon$.

Si P' es otro refinamiento de P_ε y $S(P, |f|, g)$ es una suma de Riemann-Stieltjes de $|f|$ con respecto a g, se tiene:

$|S(P, |f|, g) - S(P', |f|, g)|$

$\leq |S(P, |f|, g) - S(P_\varepsilon, |f|, g)| + |S(P', |f|, g) - S(P_\varepsilon, |f|, g)| < \varepsilon$.

Así que $|f|$ y g satisfacen el criterio de Cauchy.

Por lo tanto, $|f|$ es integrable con respecto a g.

Finalmente, como $|f| \geq f$ y $|f| \geq -f$, entonces:

$\int_a^b |f|\, dg \geq \int_a^b f dg,$

$\int_a^b |f|\, dg \geq -\int_a^b f dg.$

Por lo tanto:

$\int_a^b |f|\, dg \geq \left|\int_a^b f dg\right|.$

■

El siguiente resultado es importante ya que muestra que la propiedad de que la función integradora sea de variación acotada se conserva para la integral que se obtiene. Esto hace que la integral de cualquier función continua con respecto a la función que se obtiene al integrar esté bien definida.

TEOREMA 4.6. *Sean $f : [a, b] \to \mathbb{R}$ una función continua y $g : [a, b] \to \mathbb{R}$ una función de variación acotada, entonces la función $F : [a, b] \to \mathbb{R}$ definida por:*

$F(t) = \int_a^t f dg$

es de variación acotada.

Demostración

Sean $f_1 : [a, b] \to \mathbb{R}$ y $f_2 : [a, b] \to \mathbb{R}$ dos funciones no decrecientes tales que $g = f_1 - f_2$, $M = \sup \{|f(x)| : x \in [a, b]\}$ y $P = \{x_0, x_1, \ldots, x_n\}$ una partición del intervalo $[a, b]$. Entonces:

$\sum_{k=1}^n |F(x_k) - F(x_{k-1})| = \sum_{k=1}^n \left|\int_{x_{k-1}}^{x_k} f dg\right| \leq \sum_{k=1}^n \int_{x_{k-1}}^{x_k} |f|\, df_1 + \sum_{k=1}^n \int_{x_{k-1}}^{x_k} |f|\, df_2$

$= \int_a^b |f|\, df_1 + \int_a^b |f|\, df_2 \leq M [f_1(b) - f_1(a)] + M [f_2(b) - f_2(a)].$

Así que:

$V_F[a, b] = \sup \{V_g(P) : P \text{ es una partición de } [a, b]\}$

$\leq M [f_1(b) - f_1(a)] + M [f_2(b) - f_2(a)] < \infty.$

■

TEOREMA 4.7. *Sean $f : [a, b] \to \mathbb{R}$ una función continua y $g : [a, b] \to \mathbb{R}$ una función continua de variación acotada, entonces la función $F : [a, b] \to \mathbb{R}$ definida por:*

$F(t) = \int_a^t f dg$

es continua.

Demostración

Sean $f_1 : [a,b] \to \mathbb{R}$ y $f_2 : [a,b] \to \mathbb{R}$ dos funciones continuas no decrecientes tales que $g = f_1 - f_2$, $M_1 = \sup\{|f_1(x)| : x \in [a,b]\} + 1$, $M_2 = \sup\{|f_2(x)| : x \in [a,b]\} + 1$ y $M = \sup\{|f(x)| : x \in [a,b]\} + 1$.

Dada $\varepsilon > 0$, sea $h \in (0, b-a)$ tal que:

$f_1(a+h) - f_1(a) < \frac{1}{2M_1}\varepsilon$,

$f_2(a+h) - f_2(a) < \frac{1}{2M_2}\varepsilon$,

$f_1(b) - f_1(b-h) < \frac{1}{2M_1}\varepsilon$,

$f_2(b) - f_2(b-h) < \frac{1}{2M_2}\varepsilon$.

Entonces:

$\left|\int_a^{a+h} f df_1\right| \le \int_a^{a+h} |f|\, df_1 \le M_1 [f_1(a+h) - f_1(a)] < \frac{1}{2}\varepsilon$,

$\left|\int_a^{a+h} f df_2\right| \le \int_a^{a+h} |f|\, df_2 \le M_2 [f_1(a+h) - f_1(a)] < \frac{1}{2}\varepsilon$,

$\left|\int_{b-h}^{b} f df_1\right| \le \int_{b-h}^{b} |f|\, df_1 \le M_1 [f_1(b) - f_1(b-h)] < \frac{1}{2}\varepsilon$,

$\left|\int_{b-h}^{b} f df_2\right| \le \int_{b-h}^{b} |f|\, df_2 \le M_2 [f_2(b) - f_2(b-h)] < \frac{1}{2}\varepsilon$.

Así que:

$|F(a+h) - F(a)| = \left|\int_a^{a+h} f df_1 - \int_a^{a+h} f df_2\right| \le \left|\int_a^{a+h} f df_1\right| + \left|\int_a^{a+h} f df_2\right| < \varepsilon$,

$|F(b) - F(b-h)| = \left|\int_{b-h}^{b} f df_1 - \int_{b-h}^{b} f df_2\right| \le \left|\int_{b-h}^{b} f df_1\right| + \left|\int_{b-h}^{b} f df_2\right| < \varepsilon\}$.

Por lo tanto, F es continua por la derecha en a y continua por la izquierda en b.

Si $u \in (a,b)$, dada $\varepsilon > 0$, sea $h \in (0, \text{mín}(u-a, b-u))$ tal que:

$|f_1(x) - f_1(u)| < \frac{1}{2M}\varepsilon$,

$|f_2(x) - f_2(u)| < \frac{1}{2M}\varepsilon$

para cualquier $x \in (u-h, u+h)$.

Entonces:

$|F(x) - F(u)| = \left|\int_{\text{mín}(x,u)}^{\text{máx}(x,u)} f dg\right| \le \int_{\text{mín}(x,u)}^{\text{máx}(x,u)} |f|\, df_1 + \int_{\text{mín}(x,u)}^{\text{máx}(x,u)} |f|\, df_2$

$\leq M\left[f_1\left(\max\left(x,u\right)\right)-f_1\left(\min\left(x,u\right)\right)\right]+M\left[f_2\left(\max\left(x,u\right)\right)-f_2\left(\min\left(x,u\right)\right)\right]<\varepsilon.$

Así que F es continua en u.

■

PROPOSICIÓN 4.4. *Sean $f:[a,b]\to\mathbb{R}$ una función continua y $g:[a,b]\to\mathbb{R}$ una función no decreciente, entonces existe $c\in[a,b]$ tal que:*

$\int_a^b fdg=f\left(c\right)\left[g\left(b\right)-g\left(a\right)\right].$

Demostración

Si g es constante, el resultado es inmediato, así que supongamos que g no es constante.

Sea $M=\sup\left\{f\left(x\right):x\in[a,b]\right\}$ y $m=\inf\left\{f\left(x\right):x\in[a,b]\right\}$, entonces:

$\int_a^b mdg\leq\int_a^b fdg\leq\int_a^b Mdg.$

Así que:

$m\leq\frac{1}{g(b)-g(a)}\int_a^b fdg\leq M.$

Como f es continua, existe $c\in[a,b]$ tal que:

$f\left(c\right)=\frac{1}{g(b)-g(a)}\int_a^b fdg.$

■

TEOREMA 4.8. *Sean $f:[a,b]\to\mathbb{R}$ y $h:[a,b]\to\mathbb{R}$ dos funciones continuas, $g:[a,b]\to\mathbb{R}$ una función de variación acotada y $F:[a,b]\to\mathbb{R}$ definida por:*

$F\left(t\right)=\int_a^t fdg.$

Entonces:

$\int_a^b hdF=\int_a^b hfdg.$

Demostración

Sean $f_1:[a,b]\to\mathbb{R}$ y $f_2:[a,b]\to\mathbb{R}$ dos funciones no decrecientes tales que $g=f_1-f_2$, $M=\sup\left\{\left|h\left(x\right)\right|:x\in[a,b]\right\}+1$, $c_1=f_1\left(b\right)-f_1\left(a\right)+1$, $c_2=f_2\left(b\right)-f_2\left(a\right)+1$, $c=\max\left(c_1,c_2\right)$ y, dada $\varepsilon>0$, sea $\delta>0$ tal que, si $x,y\in[a,b]$ y $\left|y-x\right|<\delta$, entonces:

$\left|f\left(y\right)-f\left(x\right)\right|<\frac{1}{4cM}\varepsilon.$

$\left|h(y)f\left(y\right)-h(x)f\left(x\right)\right|<\frac{1}{4c}\varepsilon.$

Sea P_ε una partición del intervalo $[a,b]$ de norma menor que δ, $P=\left\{x_0,x_1,\ldots,x_n\right\}$ un refinamiento de P_ε, $S(P,h,F)=\sum_{k=1}^n h(\xi_k)\left[F(x_k)-F(x_{k-1})\right]$ una suma de Riemann-Stieltjes de h con respecto a y, para cada $k\in\{1,2,\ldots n\}$, $\zeta_k^{(1)},\zeta_k^{(2)},\eta_k^{(1)},\eta_k^{(2)}\in[x_{k-1},x_k]$ tales que:

$\int_{x_{k-1}}^{x_k} f df_1 = f\left(\zeta_k^{(1)}\right)[f_1(x_k) - f_1(x_{k-1})]$,

$\int_{x_{k-1}}^{x_k} f df_2 = f\left(\zeta_k^{(2)}\right)[f_2(x_k) - f_2(x_{k-1})]$,

$\int_{x_{k-1}}^{x_k} hf df_1 = h(\eta_k^{(1)}) f\left(\eta_k^{(1)}\right)[f_1(x_k) - f_1(x_{k-1})]$,

$\int_{x_{k-1}}^{x_k} hf df_2 = h(\eta_k^{(2)}) f\left(\eta_k^{(2)}\right)[f_2(x_k) - f_2(x_{k-1})]$.

Entonces:

$\left|S(P,h,F) - \int_a^b hfdg\right| = \left|\sum_{k=1}^n h(\xi_k)[F(x_k) - F(x_{k-1})] - \sum_{k=1}^n \int_{x_{k-1}}^{x_k} hfdg\right|$

$= \left|\sum_{k=1}^n \left[h(\xi_k)\int_{x_{k-1}}^{x_k} fdg - \int_{x_{k-1}}^{x_k} hfdg\right]\right|$

$= \left|\sum_{k=1}^n \left[h(\xi_k)\int_{x_{k-1}}^{x_k} fdf_1 - \int_{x_{k-1}}^{x_k} hfdf_1\right] - \sum_{k=1}^n \left[h(\xi_k)\int_{x_{k-1}}^{x_k} fdf_2 - \int_{x_{k-1}}^{x_k} hfdf_2\right]\right|$

$\leq \left|\sum_{k=1}^n \left[h(\xi_k)\int_{x_{k-1}}^{x_k} fdf_1 - \int_{x_{k-1}}^{x_k} hfdf_1\right]\right| + \left|\sum_{k=1}^n \left[h(\xi_k)\int_{x_{k-1}}^{x_k} fdf_2 - \int_{x_{k-1}}^{x_k} hfdf_2\right]\right|$

$= \left|\sum_{k=1}^n \left[h(\xi_k)f\left(\zeta_k^{(1)}\right)(f_1(x_k) - f_1(x_{k-1})) - h(\eta_k^{(1)})f\left(\eta_k^{(1)}\right)(f_1(x_k) - f_1(x_{k-1}))\right]\right|$

$+ \left|\sum_{k=1}^n \left[h(\xi_k)f\left(\zeta_k^{(2)}\right)(f_2(x_k) - f_2(x_{k-1})) - h(\eta_k^{(2)})f\left(\eta_k^{(2)}\right)(f_2(x_k) - f_2(x_{k-1}))\right]\right|$

$= \left|\sum_{k=1}^n \left[h(\xi_k)f\left(\zeta_k^{(1)}\right) - h(\eta_k^{(1)})f\left(\eta_k^{(1)}\right)\right][f_1(x_k) - f_1(x_{k-1})]\right|$

$+ \left|\sum_{k=1}^n \left[h(\xi_k)f\left(\zeta_k^{(2)}\right) - h(\eta_k^{(2)})f\left(\eta_k^{(2)}\right)\right][f_2(x_k) - f_2(x_{k-1})]\right|$

$\leq \sum_{k=1}^n \left|h(\xi_k)f\left(\zeta_k^{(1)}\right) - h(\eta_k^{(1)})f\left(\eta_k^{(1)}\right)\right|[f_1(x_k) - f_1(x_{k-1})]$

$+ \sum_{k=1}^n \left|h(\xi_k)f\left(\zeta_k^{(2)}\right) - h(\eta_k^{(2)})f\left(\eta_k^{(2)}\right)\right|[f_2(x_k) - f_2(x_{k-1})]$

$\leq \sum_{k=1}^n \left[\left|h(\xi_k)f(\xi_k) - h(\eta_k^{(1)})f\left(\eta_k^{(1)}\right)\right| + |h(\xi_k)|\left|f\left(\zeta_k^{(1)}\right) - f(\xi_k)\right|\right][f_1(x_k) - f_1(x_{k-1})]$

$+ \sum_{k=1}^n \left[\left|h(\xi_k)f(\xi_k) - h(\eta_k^{(2)})f\left(\eta_k^{(2)}\right)\right| + |h(\xi_k)|\left|f\left(\zeta_k^{(2)}\right) - f(\xi_k)\right|\right][f_2(x_k) - f_2(x_{k-1})]$

$\leq \left(\frac{1}{4c}\varepsilon + M\frac{1}{4cM}\varepsilon\right)\sum_{k=1}^n [f_1(x_k) - f_1(x_{k-1})] + \left(\frac{1}{4c}\varepsilon + M\frac{1}{4cM}\varepsilon\right)\sum_{k=1}^n [f_2(x_k) - f_2(x_{k-1})]$

$= \left(\frac{1}{4c}\varepsilon + \frac{1}{4c}\varepsilon\right)[f_1(b) - f_1(a)] + \left(\frac{1}{4c}\varepsilon + \frac{1}{4c}\varepsilon\right)[f_2(b) - f_2(a)]$

$< \left(\frac{1}{2c}\varepsilon\right)c + \left(\frac{1}{2c}\varepsilon\right)c = \varepsilon.$

■

Al igual que la integral de Riemann, la integral de Riemann Stieltjes presenta limitaciones en cuanto a la convergencia de las integrales de una sucesión convergente de funciones. El resultado siguiente muestra que el límite y la integral se pueden intercambiar pero la condición de convergencia uniforme que se pide es muy fuerte. Es éste uno de los aspectos en que la integral de Lebesgue y la integral de Lebesgue-Stieltjes, que definiremos posteriormente, superan a la integral de Riemann y a la de Riemann-Stieltjes, respectivamente.

TEOREMA 4.9. *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones $f_n : [a,b] \to \mathbb{R}$, integrables con respecto a una función no decreciente $g : [a,b] \to \mathbb{R}$, y supongamos que $(f_n)_{n\in\mathbb{N}}$ converge uniformemente a la función $f : [a,b] \to \mathbb{R}$. Entonces, f es integrable con respecto a g y:*

$\int_a^b f dg = \lim_{n\rightsquigarrow\infty} \int_a^b f_n dg.$

Demostración

Dada $\varepsilon > 0$, sea $N \in \mathbb{N}$ tal que:

$|f_n(x) - f(x)| < \frac{1}{2[g(b)-g(a)]+1}\varepsilon$

para cualquier número natural $n \geq N$ y cualquier $x \in [a,b]$, y sea P_ε una partición del intervalo $[a,b]$ tal que si $P = \{x_0, x_1, \ldots, x_n\}$ y P' son dos refinamientos de P_ε y $S(P, f_N, g)$, $S(P', f_N, g)$ son sumas de Riemann-Stieltjes de f_N con respecto a g, entonces:

$|S(P, f_N, g) - S(P', f_N, g)| < \frac{1}{2[g(b)-g(a)]+1}\varepsilon.$

Sean $S(P, f, g) = \sum_{k=1}^n f(\xi_k)[g(x_k) - g(x_{k-1})]$ y $S(P', f, g) = \sum_{k=1}^m f(\zeta_k)[g(y_k) - g(y_{k-1})]$ dos sumas de Riemann-Stieltjes de f con respecto a g y definamos:

$M = \sup\{|f_N(x) - f(x)| : x \in [a,b]\},$

$S(P, f_N, g) = \sum_{k=1}^n f_N(\xi_k)[g(x_k) - g(x_{k-1})],$

$S(P', f_N, g) = \sum_{k=1}^m f_N(\zeta_k)[g(y_k) - g(y_{k-1})].$

Se tiene entonces:

$|S(P, f, g) - S(P, f_N, g)| = |\sum_{k=1}^n f(\xi_k)[g(x_k) - g(x_{k-1})] - \sum_{k=1}^n f_N(\xi_k)[g(x_k) - g(x_{k-1})]|$

$\leq M \sum_{k=1}^n |g(x_k) - g(x_{k-1})| \leq \frac{g(b)-g(a)}{2[g(b)-g(a)]+1}\varepsilon.$

$|S(P', f_N, g) - S(P', f, g)| = |\sum_{k=1}^m f_N(\zeta_k)[g(y_k) - g(y_{k-1})] - \sum_{k=1}^m f(\zeta_k)[g(y_k) - g(y_{k-1})]|$

$\leq M \sum_{k=1}^m |g(y_k) - g(y_{k-1})| \leq \frac{g(b)-g(a)}{2[g(b)-g(a)]+1}\varepsilon.$

Por lo tanto:

$|S(P, f, g) - S(P', f, g)|$

$\leq |S(P, f, g) - S(P, f_N, g)| + |S(P, f_N, g) - S(P', f_N, g)| + |S(P', f_N, g) - S(P', f, g)|$

$< \frac{g(b)-g(a)}{2[g(b)-g(a)]+1}\varepsilon + \frac{1}{2[g(b)-g(a)]+1}\varepsilon + \frac{g(b)-g(a)}{2[g(b)-g(a)]+1}\varepsilon = \varepsilon.$

Así que f y g satisfacen el criterio de Cauchy, de lo cual se sigue que f es integrable con respecto a g.

Finalmente, para cualquier $n \geq N$, se tiene:

$\left|\int_a^b f_n dg - \int_a^b f dg\right| = \left|\int_a^b (f_n - f)\, dg\right| \leq \int_a^b |f_n - f|\, dg \leq \frac{g(b)-g(a)}{2[g(b)-g(a)]+1}\varepsilon \leq \varepsilon.$

Por lo tanto:

$\lim_{n\rightsquigarrow\infty} \int_a^b f_n dg = \int_a^b f dg.$

■

Corolario 4.2. *Sean $g : [a, b] \to \mathbb{R}$ una función de variación acotada y $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones continuas $f_n : [a, b] \to \mathbb{R}$ que converge uniformemente a la función $f : [a, b] \to \mathbb{R}$. Entonces:*

$\int_a^b f dg = \lim_{n\rightsquigarrow\infty} \int_a^b f_n dg.$

Demostración

Sean $g_1 : [a, b] \to \mathbb{R}$ y $g_2 : [a, b] \to \mathbb{R}$ dos funciones no decrecientes tales que $g = g_1 - g_2$. Por la proposición 4.9, se tiene:

$\int_a^b f dg_1 = \lim_{n\rightsquigarrow\infty} \int_a^b f_n dg_1,$

$\int_a^b f dg_2 = \lim_{n\rightsquigarrow\infty} \int_a^b f_n dg_2.$

Así que:

$\int_a^b f dg = \int_a^b f dg_1 - \int_a^b f dg_2 = \lim_{n\to\infty} \int_a^b f_n dg_1 - \lim_{n\to\infty} \int_a^b f_n dg_2 = \lim_{n\rightsquigarrow\infty} \int_a^b f_n dg.$

■

4.3. Funciones de variación acotada e integrabilidad de las funciones continuas

Lo siguiente tiene como objetivo demostrar que si $g : [a, b] \to \mathbb{R}$ no es de variación acotada, entonces existe una función continua $f : [a, b] \to \mathbb{R}$ tal que f no es integrable con respecto a g, lo cual tiene como corolario que si toda función continua $f : [a, b] \to \mathbb{R}$ es integrable con respecto a una función $g : [a, b] \to \mathbb{R}$, entonces g es de variación acotada.

Definición 4.4. *Diremos que una función $g : [a, b] \to \mathbb{R}$ es de variación acotada en un punto $x_0 \in (a, b)$ si existe $\delta > 0$ tal que $[x_0 - \delta, x_0 + \delta] \subset [a, b]$ y g es de variación acotada* *en* $[x_0 - \delta, x_0 + \delta]$*. Diremos que* g *es de variación acotada en* a *(resp.* b*) si existe* $\delta > 0$ *tal que* $[a, a + \delta] \subset [a, b]$ *(resp.* $[b - \delta, b] \subset [a, b]$*) y* g *es de variación acotada en* $[a, a + \delta]$ *(resp.* $[b - \delta, b]$*).*

Diremos que g *es localmente de variación acotada en* $[a, b]$ *si es de variación acotada en cada punto* $x_0 \in [a, b]$*.*

Teorema 4.10. *Una función* $g : [a, b] \to \mathbb{R}$ *es de variación acotada en* $[a, b]$ *si y sólo si es localmente de variación acotada en* $[a, b]$*.*

Demostración

Si g es de variación acotada y $x_0 \in (a, b)$, tomando cualquier $\delta > 0$ tal que $[x_0 - \delta, x_0 + \delta] \subset [a, b]$, g es de variación acotada en $[x_0 - \delta, x_0 + \delta]$; de la misma manera, tomando cualquier $\delta > 0$ tal que $[a, a + \delta] \subset [a, b]$ (resp. $[b - \delta, b] \subset [a, b]$), g es de variación acotada en $[a, a + \delta]$ (resp. $[b - \delta, b]$). Así que g es localmente de variación acotada.

Supongamos ahora que g es localmente de variación acotada en $[a, b]$.

Para cada $x \in (a, b)$, sea $\delta_x > 0$ tal que $[x - \delta_x, x + \delta_x] \subset [a, b]$ y g es de variación acotada en $[x - \delta_x, x + \delta_x]$. Sea también $\delta_a > 0$ (resp. $\delta_b > 0$) tal que $[a, a + \delta_a] \subset [a, b]$ (resp. $[b - \delta_b, b] \subset [a, b]$) y g es de variación acotada en $[a, a + \delta_a]$ (resp. $[b - \delta_b, b]$).

Tomando $\delta > 0$ arbitraria, la colección de intervalos

$(a - \delta, a + \delta_a)$, $(b - \delta_b, b + \delta)$ y $(x - \delta_x, x + \delta_x)$,

para cada $x \in (a, b)$, constituye una cubierta de $[a, b]$ con intervalos abiertos; por el teorema de Heine-Borel existe entonces una subcubierta finita $(a - \delta, a + \delta_a)$, $(b - \delta_b, b + \delta)$, $(x_1 - \delta_{x_1}, x_1 + \delta_{x_1})$, $(x_2 - \delta_{x_2}, x_2 + \delta_{x_2})$, $\ldots$, $(x_n - \delta_{x_n}, x_n + \delta_{x_n})$. Tomando a, b y los extremos de los intervalos de esta subcubierta finita (excepto $a - \delta$ y $b + \delta$) formamos una partición $P = \{a = y_0 < y_1 < \cdots < y_m = b\}$ de $[a, b]$. Obviamente cada subintervalo (y_{k-1}, y_k), con $k \in \{1, 2, \ldots, n\}$, está contenido en algún intervalo de la subcubierta finita, por lo tanto, g es de variación acotada en cada uno de los intervalos $[y_{k-1}, y_k]$ y, entonces, es de variación acotada en $[a, b]$.

■

Proposición 4.5. *Sea* $g : [a, b] \to \mathbb{R}$ *una función acotada que no es de variación acotada en* $[a, b]$*. Entonces existe* $w \in [a, b]$ *para el cual se cumple alguna de las dos condiciones siguientes:*

a) Existe una sucesión decreciente $(y_n)_{n \in \{0,1,\ldots\}}$ *de números reales en* $[a, b]$ *tal que:*

i) $y_0 = b$*.*

ii) $\lim_{n \rightsquigarrow \infty} y_n = w$*.*

iii) $\sum_{n=1}^{\infty} |g(y_{n-1}) - g(y_n)| = \infty$*.*

b) Existe una sucesión creciente $(x_n)_{n\in\{0,1,\ldots\}}$ *de números reales en* $[a,b]$ *tal que:*

i) $x_0 = a$.

ii) $\lim_{n\rightsquigarrow\infty} x_n = w$.

iii) $\sum_{n=1}^{\infty} |g(x_n) - g(x_{n-1})| = \infty$.

Demostración

Como g no es de variación acotada en $[a,b]$, existe $w \in [a,b]$ tal que g no es de variación acotada en w.

Si $w = a$, entonces, dada cualquier $\delta \in (0, b-a]$, g no es de variación acotada en $[a, a+\delta]$.

Si $w = b$, entonces dada cualquier $\delta \in (0, b-a]$, g no es de variación acotada en $[b-\delta, b]$.

Si $w \in (a,b)$ y existe $\delta_0 \in (0, w-a]$ tal que g es de variación acotada en $[w-\delta_0, w]$, entonces dada cualquier $\delta \in (0, b-w]$, g no es de variación acotada en $[w, w+\delta]$.

Si $w \in (a,b)$ y existe $\delta_0 \in (0, b-w]$ tal que g es de variación acotada en $[w, w+\delta_0]$, entonces dada cualquier $\delta \in (0, w-a]$, g no es de variación acotada en $[w-\delta, w]$.

Los cuatro casos anteriores se pueden reducir a los dos siguientes:

1) $w \in [a,b)$ y dada cualquier $\delta \in (0, b-w]$, g no es de variación acotada en $[w, w+\delta]$.

2) $w \in (a,b]$ y dada cualquier $\delta \in (0, w-a]$, g no es de variación acotada en $[w-\delta, w]$.

En el primer caso se cumple la condición a de la proposición; en el segundo se cumple la condición b. Las demostraciones de estas dos aseveraciones son similares; demostremos la segunda.

Supongamos entonces que $w \in (a,b]$ y que dada cualquier $\delta \in (0, w-a]$, g no es de variación acotada en $[w-\delta, w]$.

Vamos a demostrar primero que existe una sucesión creciente $(z_n)_{n\in\{0,1,\ldots\}}$ de números reales en $[a,b]$ tal que:

i) $z_0 = a$.

ii) $\lim_{n\rightsquigarrow\infty} z_n = w$.

iii) Para cada $n \in \{0,1,\ldots\}$, existe una partición P_n del intervalo $[z_n, z_{n+1}]$ tal que $V_g(P_n) > 1$.

Definamos $M = \sup\{|g(x)| : x \in [a,b]\}$ y $z_0 = a$.

Paso 0. Como g no es de variación acotada en $[z_0, w]$, existe una partición:

$Q_0 = \{z_0 = u_{0,0} < u_{0,1} < \cdots < u_{0,m_0} = w\}$

tal que:

$V_g(Q_0) > 2M + 1$ y $w - u_{0,m_0-1} < \frac{1}{2}$.

Definamos $z_1 = u_{0,m_0-1}$ y llamemos P_0 a la partición Q_0 restringida al intervalo $[z_0, z_1]$.

Si $V_g(P_0) \leq 1$, se tendría:

$V_g(Q_0) = V_g(P_0) + |g(w) - g(z_1)| \leq 1 + 2M$,

lo cual es una contradicción.

Por lo tanto, $V_g(P_0) > 1$.

Paso 1. Como g no es de variación acotada en $[z_1, w]$, existe una partición:

$Q_1 = \{z_1 = u_{1,0} < u_{1,1} < \cdots < u_{1,m_1} = w\}$

tal que:

$V_g(Q_1) > 2M + 1$ y $w - u_{1,m_1-1} < \frac{1}{2^2}$.

Definamos $z_2 = u_{1,m_1-1}$ y llamemos P_1 a la partición Q_1 restringida al intervalo $[z_1, z_2]$.

Si $V_g(P_1) \leq 1$, se tendría:

$V_g(Q_1) = V_g(P_1) + |g(w) - g(z_2)| \leq 1 + 2M$,

lo cual es una contradicción.

Por lo tanto, $V_g(P_1) > 1$.

Continuando con este procedimiento hasta el paso k, tendremos definido el conjunto

$\{z_0, z_1, \ldots, z_{k+1}\}$

de tal forma que:

$a = z_0 < z_1 < \cdots < z_{k+1} < w$,

Para $j \in \{1, 2, \ldots, k+1\}$, $w - z_j < \frac{1}{2^j}$,

Para cada $j \in \{0, 1, 2, \ldots, k\}$, existe una partición P_j del intervalo $[z_j, z_{j+1}]$ tal que $V_g(P_j) > 1$,

Paso $k+1$. Como g no es de variación acotada en $[z_{k+1}, w]$, existe una partición:

$Q_{k+1} = \left\{z_{k+1} = u_{k+1,0} < u_{k+1,1} < \cdots < u_{k+1,m_{k+1}} = w\right\}$

tal que:

$V_g(Q_{k+1}) > 2M + 1$ y $w - u_{k+1,m_{k+1}-1} < \frac{1}{2^{k+2}}$.

Definamos $z_{k+2} = u_{k+1,m_{k+1}-1}$ y llamemos P_{k+1} a la partición Q_{k+1} restringida al intervalo $[z_{k+1}, z_{k+2}]$.

Si $V_g(P_{k+1}) \leq 1$, se tendría:

$V_g(Q_{k+1}) = V_g(P_{k+1}) + |g(w) - g(z_{k+2})| \leq 1 + 2M$,

lo cual es una contradicción.

Por lo tanto, $V_g(P_{k+1}) > 1$.

Por el principio de inducción matemática, para cada $n \in \{0, 1, \ldots\}$, existe $z_{n+1} \in [a, b]$ tal que:

$a = z_0 < z_1 < \cdots < z_{n+1} < w$,

$w - z_{n+1} < \frac{1}{2^{n+1}}$,

Existe una partición P_n del intervalo $[z_n, z_{n+1}]$ tal que $V_g(P_n) > 1$.

Los puntos de la unión $\cup_{n=0}^{\infty} P_n$, ordenados en forma creciente, constituyen una sucesión $(x_n)_{n\in\{0,1,\ldots\}}$ de números reales en $[a, b]$ tal que $x_0 = a$, $\lim_{n\rightsquigarrow\infty} x_n = w$ y:

$\sum_{n=1}^{\infty} |g(x_n) - g(x_{n-1})| = \sum_{n=0}^{\infty} V_g(P_n) = \infty$.

■

LEMA 4.1. *Sea $(a_n)_{n\in\mathbb{N}}$ una sucesión de números reales no negativos tales que $\sum_{n=1}^{\infty} a_n = \infty$. Entonces existe una sucesión no creciente $(c_n)_{n\in\mathbb{N}}$ de números reales positivos tales que $\lim_{n\to\infty} c_n = 0$ y $\sum_{n=1}^{\infty} c_n a_n = \infty$.*

Demostración

Para cada $n \in \mathbb{N}$, definamos $s_n = \sum_{k=1}^{n} a_k$.

Sea a_N el primer elemento positivo de la sucesión $(a_n)_{n\in\mathbb{N}}$, entonces, para $k \in \{1, \ldots, N\}$ definamos $c_k = \frac{1}{a_N}$ y, para $k \in \{N+1, N+2, \ldots\}$, $c_k = \frac{1}{s_k}$.

También, para cada $n \in \mathbb{N}$, definamos $t_n = \sum_{k=1}^{n} c_k a_k$.

Si $n, m \in \mathbb{N}$ y $n > m$, se tiene:

$t_n - t_m = \sum_{k=m+1}^{n} c_k a_k \geq \sum_{k=m+1}^{n} c_n a_k = c_n (s_n - s_m) = 1 - \frac{s_m}{s_n}$.

Como $\lim_{n\rightsquigarrow\infty} s_n = \infty$, fijando m podemos tomar n suficientemente grande de tal manera que $s_n > 2s_m$, en cuyo caso se tiene:

$t_n - t_m \geq 1 - \frac{s_m}{s_n} > \frac{1}{2}$.

En otras palabras, dado $m \in \mathbb{N}$, existe $n \in \mathbb{N}$ tal que $n > m$ y $t_n - t_m > \frac{1}{2}$.

Por lo tanto, la sucesión no decreciente $(t_n)_{n\in\mathbb{N}}$ no es de Cauchy, así que no converge y entonces:

$\sum_{n=1}^{\infty} c_n a_n = \lim_{n\rightsquigarrow\infty} t_n = \infty$.

■

Teorema 4.11. *Sea $g : [a, b] \to \mathbb{R}$ una función acotada que no es de variación acotada en $[a, b]$. Entonces existe una función continua $f : [a, b] \to \mathbb{R}$ la cual no es Riemann-Stieltjes integrable con respecto a g.*

Demostración

Sabemos que existe $w \in [a, b]$ para el cual se cumple alguna de las dos condiciones de la proposición anterior; supongamos que se cumple la segunda (así que $w \in (a, b]$) y consideremos una sucesión creciente $(x_n)_{n\in\{0,1,\dots\}}$ de números reales en $[a, b]$ tal que $x_0 = a$, $\lim_{n\to\infty} x_n = w$ y $\sum_{n=1}^{\infty} |g(x_n) - g(x_{n-1})| = \infty$.

Consideremos también una sucesión no creciente $(c_n)_{n\in\mathbb{N}}$ de números reales positivos tales que $\lim_{n\rightsquigarrow\infty} c_n = 0$ y:

$\sum_{n=1}^{\infty} c_n |g(x_n) - g(x_{n-1})| = \infty$.

Definamos $f : [a, b] \to \mathbb{R}$ de la siguiente manera:

Para $x \in \left[a, \frac{x_0+x_1}{2}\right]$, $f(x) = a_0 x + b_0$, de tal manera que:

$f(a) = 0$,

$f(\frac{x_0+x_1}{2}) = \begin{cases} c_1 & \text{si } g(x_1) - g(x_0) \geq 0 \\ -c_1 & \text{si } g(x_1) - g(x_0) < 0 \end{cases}$

Para cada $k \in \mathbb{N}$ y $x \in \left[\frac{x_{k-1}+x_k}{2}, \frac{x_k+x_{k+1}}{2}\right]$, $f(x) = a_k x + b_k$, de tal manera que:

$f(\frac{x_{k-1}+x_k}{2}) = \begin{cases} c_k & \text{si } g(x_k) - g(x_{k-1}) \geq 0 \\ -c_k & \text{si } g(x_k) - g(x_{k-1}) < 0 \end{cases}$

$f(\frac{x_k+x_{k+1}}{2}) = \begin{cases} c_{k+1} & \text{si } g(x_{k+1}) - g(x_k) \geq 0 \\ -c_{k+1} & \text{si } g(x_{k+1}) - g(x_k) < 0 \end{cases}$

Para $x \in [w, b]$, $f(x) = 0$.

Dada una partición $P = \{v_0, v_1, \dots, v_N\}$ de $[a, b]$ y $M > 0$, definamos:

$r = máx \{j \in \{0, 1, 2, \dots, N\} : v_j < w\}$,

$s = mín\{j \in \{1, 2, \ldots\} : x_j > v_r\}$,

$S = \sum_{j=1}^{r} f(v_{j-1})\,[g(v_j) - g(v_{j-1})] + f(v_r)\,[g(x_s) - g(v_r)]$,

$k_M \in \mathbb{N}$ tal que $\sum_{n=s+1}^{k_M} c_n\,|g(x_n) - g(x_{n-1})| > M - S$,

$P' = \{v_0, \ldots, v_r, x_s, \ldots, x_{k_M}, w, v_{r+1}, \ldots, v_N\}$,

$S(P', f, g) = \sum_{j=1}^{r} f(v_{j-1})\,[g(v_j) - g(v_{j-1})] + f(v_r)\,[g(x_s) - g(v_r)]$

$+ \sum_{j=s+1}^{k_M} f(\frac{x_{j-1}+x_j}{2})\,[g(x_j) - g(x_{j-1})] + f(w)\,[g(w) - g(x_{k_M})]$

$+f(w)\,[g(v_{r+1}) - g(w)] + \sum_{j=r+1}^{N-1} f(v_j)\,[g(v_{j+1}) - g(v_j)]$

donde una sumatoria $\sum_{j=i_1}^{i_2}$ se toma igual a cero cuando $i_2 < i_1$.

Entonces:

a) P' es un refinamiento de P.

b) $S(P', f, g)$ es una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la particion P'.

c) $S(P', f, g) = S + \sum_{j=s+1}^{k_M} c_j\,|g(x_j) - g(x_{j-1})| > M$.

Por lo tanto, no existe $I \in \mathbb{R}$ tal que para cualquier $\varepsilon > 0$ existe una particion P del intervalo $[a, b]$ tal que $|S(P', f, g) - I| < \varepsilon$ para cualquier particion P' que sea un refinamiento de P y cualquier suma de Riemann-Stieltjes $S(P', f, g)$ de f con respecto a g, correspondiente a la particion P'. Así que f no es integrable con respecto a g.

■

COROLARIO 4.3. *Si toda función continua $f : [a, b] \to \mathbb{R}$ es integrable con respecto a la función acotada $g : [a, b] \to \mathbb{R}$, entonces g es de variación acotada.*

4.4. Fórmula de integración por partes

La fórmula de integración por partes es importante no únicamente como fórmula técnica que permite expresar la integral de una función f con respecto a g en términos de la integral de g con respecto a f. Una parte muy importante del resultado es que si una función f es integrable con respecto a una función g, entonces g es integrable con respecto a f. Esto permite afirmar, por ejemplo, que toda función de variación acotada es integrable con respecto a cualquier función continua.

TEOREMA 4.12. *Sean $f : [a, b] \to \mathbb{R}$ y $g : [a, b] \to \mathbb{R}$ dos funciones acotadas y supongamos que f es integrable con respecto a g, entonces g es integrable con respecto a f y, además, se tiene:*

$\int_a^b gdf = g(b)f(b) - g(a)f(a) - \int_a^b fdg$.

Demostración

Dada $\varepsilon > 0$, sea P_ε una partición del intervalo $[a,b]$ tal que $\left|S(P,f,g) - \int_a^b fdg\right| < \varepsilon$ para cualquier partición P que sea un refinamiento de P_ε y cualquier suma de Riemann-Stieltjes $S(P,f,g)$ de f con respecto a g, correspondiente a la particion P. Sea entonces $P = \{x_0, x_1, \ldots, x_n\}$ un refinamiento de P_ε y consideremos ahora una suma de Riemann-Stieltjes $S(P,g,f)$ de g con respecto a f, correspondiente a la particion P:

$$S(P,g,f) = \sum_{k=1}^n g(\xi_k)\,[f(x_k) - f(x_{k-1})],$$

la cual tambien puede escribirse en la forma siguiente

$$S(P,g,f) = \sum_{k=1}^n g(\xi_k)\,[f(x_k) - f(x_{k-1})]$$

$$= \sum_{k=1}^{n-1} f(x_k)\left[g(\xi_k) - g(\xi_{k+1})\right] + f(x_n)g(\xi_n) - f(x_0)g(\xi_1)$$

$$= \sum_{k=1}^{n-1} f(x_k)\left[g(\xi_k) - g(x_k) + g(x_k) - g(\xi_{k+1})\right] + f(x_n)g(\xi_n) - f(x_0)g(\xi_1)$$

$$= \sum_{k=1}^{n-1}\left(f(x_k)\,[g(\xi_k) - g(x_k)] + f(x_k)\left[g(x_k) - g(\xi_{k+1})\right]\right) + f(x_n)g(\xi_n) - f(x_0)g(\xi_1)$$

$$= f(x_n)g(x_n) - f(x_0)g(x_0) - f(x_0)\,[g(\xi_1) - g(x_0)]$$

$$- \sum_{k=1}^{n-1}\left(f(x_k)\,[g(x_k) - g(\xi_k)] + f(x_k)\left[g(\xi_{k+1}) - g(x_k)\right]\right) - f(x_n)\,[g(x_n) - g(\xi_n)]$$

$$= f(x_n)g(x_n) - f(x_0)g(x_0) - S(P',f,g) = f(b)g(b) - f(a)g(a) - S(P',f,g),$$

donde $P' = \{a = x_0 \le \xi_1 \le x_1 \le \xi_2 \le x_2 \le \cdots \le x_{n-1} \le \xi_n \le x_n = b\}$.

Como P' es un refinamiento de P_ε, se tiene $\left|S(P',f,g) - \int_a^b fdg\right| < \varepsilon$; es decir:

$$\left|S(P,g,f) - \left[f(b)g(b) - f(a)g(a) - \int_a^b fdg\right]\right| < \varepsilon.$$

Por lo tanto g es integrable con respecto a f y:

$$\int_a^b gdf = f(b)g(b) - f(a)g(a) - \int_a^b fdg.$$

■

Corolario 4.4. *Si $f : [a,b] \to \mathbb{R}$ es continua, entonces toda función de variación acotada es integrable con respecto a f.*

El siguiente resultado es el equivalente al Teorema Fundamental del Cálculo que se demuestra para la integral de Riemann. Además, muestra que el espacio vectorial formado por las funciones de variación acotada en un intervalo cerrado y acotado es cerrado bajo la composición de cualquier elemento de ese espacio con una función de clase C^1. Esto extiende algunas propiedades que se demostraron en la segunda sección de este capítulo. Por ejemplo, se demostró que si $g : [a, b] \to \mathbb{R}$ es de variación acotada, entonces g^2 también lo es. Con base en el siguiente teorema podemos afirmar que, si g es continua, hay muchas otras funciones de g que siguen siendo de variación acotada: la composición de cualquier función de clase C^1 con g.

TEOREMA 4.13. *Sea $g : [a, b] \to \mathbb{R}$ una función continua de variación acotada y $F : \mathbb{R} \to \mathbb{R}$ una función de clase C^1 , entonces $F \circ g$ es de variación acotada y:*

$F(g(t)) = F(g(a)) + \int_a^t F'(g(s))dg(s)$

para cualquier $t \in [a, b]$.

Demostración

Se tiene:

$\int_a^t g^0(s)\, dg(s) = g(t) - g(a)$.

Supongamos que:

$\int_a^t g^k(s)\, dg(s) = \frac{1}{k+1} g^{k+1}(t) - \frac{1}{k+1} g^{k+1}(a)$

para cualquier $t \in [a, b]$, donde $k \in \{0, 1, 2, \ldots\}$.

Entonces:

$g^{k+1}(t) = g^{k+1}(a) + (k+1) \int_a^t g^k(s)\, dg(s)$

para cualquier $t \in [a, b]$.

En particular, g^{k+1} es continua y de variación acotada. Así que, por la fórmula de integración por partes, se tiene, para cualquier $t \in [a, b]$:

$g^{k+2}(t) = g^{k+1}(t)\, g(t) = g^{k+1}(a)\, g(a) + \int_a^t g^{k+1}(s)\, dg(s) + \int_a^t g(s)\, dg^{k+1}(s)$

$= g^{k+2}(a) + \int_a^t g^{k+1}(s)\, dg(s) + (k+1) \int_a^t g(s)\, g^k(s)\, dg(s)$

$= g^{k+2}(a) + (k+2) \int_a^t g^{k+1}(s)\, dg(s)$.

Así que, por el principio de inducción matemática:

$\int_a^t g^n(s)\, dg(s) = \frac{1}{n+1} g^{n+1}(t) - \frac{1}{n+1} g^{n+1}(a)$

para cualquier $n \in \{0, 1, 2, \ldots\}$ y cualquier $t \in [a, b]$.

Por lo tanto:

$g^n(t) = g^n(a) + n \int_0^t g^{n-1}(s)\, dg(s)$

para cualquier $n \in \mathbb{N}$ y cualquier $t \in [a, b]$.

Sea $p : \mathbb{R} \to \mathbb{R}$ un polinomio dado por $p(x) = \sum_{k=0}^{n} a_k x^k$, donde $n \in \mathbb{N}$. Entonces, por la linealidad de la integral:

$$p(g(t)) = \sum_{k=0}^{n} a_k g^k(t) = a_0 + \sum_{k=1}^{n} \left[a_k g^k(a) + k \int_0^t g^{k-1}(s)\, dg(s) \right]$$

$$= p(g(a)) + \int_a^t p'(g(s)) dg(s).$$

Sea $M = \sup \{|g(x)| : x \in [a,b]\}$.

Tomemos dos números reales c y d de tal forma que $c < -M$ y $d > M$, y definamos las funciones $F_c : \mathbb{R} \to \mathbb{R}$, $F_d : \mathbb{R} \to \mathbb{R}$ y $G : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:

$$F_c(x) = \left[\frac{F(-M)}{(c+M)^2} + \frac{(c+M)F'(-M)+2F(-M)}{(c+M)^3}(x+M) \right](x-c)^2,$$

$$F_d(x) = \left[\frac{F(M)}{(d-M)^2} + \frac{(d-M)F'(M)+2F(M)}{(d-M)^3}(x-M) \right](x-d)^2,$$

$$G(x) = \begin{cases} F_c(x) & \text{si } x \in [c,-M) \\ F(x) & \text{si } x \in [-M,M] \\ F_d(x) & \text{si } x \in (M,d] \\ 0 & \text{en otro caso} \end{cases}$$

G es de clase C^1 y nula fuera del intervalo (c,d), así que existe una sucesión $(p_n)_{n\in\mathbb{N}}$ de polinomios $p_n : \mathbb{R} \to \mathbb{R}$ tales que $(p_n)_{n\in\mathbb{N}}$ y $(p_n')_{n\in\mathbb{N}}$ convergen uniformente a G y G', respectivamente, en el intervalo (c,d).

Además, $G(x) = F(x)$ y $G'(x) = F'(x)$ para cualquier $x \in [-M,M]$.

Para cada $n \in \mathbb{N}$, se tiene:

$$p_n(g(t)) = p_n(g(a)) + \int_a^t p_n'(g(s)) dg(s).$$

Así que, tomando límites cuando $n \rightsquigarrow \infty$, se obtiene:

$$F(g(t)) = F(g(a)) + \int_a^t F'(g(s)) dg(s)$$

para cualquier $t \in [a,b]$.

■

4.5. Integración de funciones discontinuas

No únicamente las funciones continuas son integrables con respecto a una función de variación acotada.

Ejemplo 4.1. *Si $g : [a,b] \to \mathbb{R}$ es una función escalonada, $\{x_0, x_1, \ldots, x_n\}$ es una partición del intervalo $[a,b]$ tal que, para cualquier $j \in \{1,2,\ldots,n\}$, la función g es igual a una* *constante c_j en el intervalo (x_{j-1}, x_j) y $f : [a,b] \to \mathbb{R}$ es una función acotada y continua en $\{x_0, x_1, \dots, x_n\}$, entonces f y g son integrables, una con respecto a la otra, y se tiene:*

$\int_a^b gdf = \sum_{j=1}^n c_j \left[f(x_j) - f(x_{j-1})\right]$,

$\int_a^b fdg = f(a)\left[g(a+) - g(a)\right] + \sum_{j=1}^{n-1} f(x_j)\left[g(x_j+) - g(x_j-)\right] + f(b)\left[g(b) - g(b-)\right]$.

En efecto, como f es continua en $\{x_0, x_1, \dots, x_n\}$, si $x \in \{x_0, x_1, \dots, x_n\}$ y $h = I_{\{x\}}$, entonces h es integrable con respecto a f y $\int_a^b hdf = 0$. También, si $h_j = I_{(x_{j-1}, x_j)}$, donde $j \in \{1, 2, \dots, n\}$, entonces h_j es integrable con respecto a f y $\int_a^b h_j df = f(x_j) - f(x_{j-1})$.

Por otra parte, $g = \sum_{j=1}^n g(x_j) I_{\{x_j\}} + \sum_{j=1}^n c_j I_{(x_{j-1}, x_j)}$, así que g es integrable con respecto a f y se tiene:

$\int_a^b gdf = \sum_{j=1}^n c_j \left[f(x_j) - f(x_{j-1})\right]$,

$\int_a^b fdg = f(b)g(b) - f(a)g(a) - \int_a^b gdf$

$= f(b)g(b) - f(a)g(a) - \sum_{j=1}^n c_j \left[f(x_j) - f(x_{j-1})\right]$

$= f(b)g(b) - f(a)g(a) - \sum_{j=1}^n c_j f(x_j) + \sum_{j=0}^{n-1} c_{j+1} f(x_j)$

$= f(a)\left(c_1 - g(a)\right) + \sum_{j=1}^{n-1} f(x_j)\left(c_{j+1} - c_j\right) + f(b)\left(g(b) - c_n\right)$

$= f(a)\left[g(a+) - g(a)\right] + \sum_{j=1}^{n-1} f(x_j)\left[g(x_j+) - g(x_j-)\right] + f(b)\left[g(b) - g(b-)\right]$.

PROPOSICIÓN 4.6. *Si $g : [a,b] \to \mathbb{R}$ es una función no decreciente, continua por la derecha y que crece únicamente mediante saltos y $f : [a,b] \to \mathbb{R}$ es una función acotada y continua por la izquierda en los puntos donde g es discontinua, entonces f es integrable con respecto a g y se tiene $\int_a^b fdg = \sum_{\{x \in D\}} f(x)\left[g(x) - g(x-)\right]$, donde D es el conjunto de puntos en el intervalo $(a,b]$ donde g es discontinua.*

Demostración

Si g es constante, el resultado es trivial ya que si $g = c$, entonces $\int_a^b fdg = 0$.

Supongamos entonces que g no es constante.

Sea $D = \{x_1, x_2, \dots\}$ el conjunto de puntos en el intervalo $(a,b]$ donde g es discontinua y M una cota positiva se $|f|$.

Como $\sum_{\{x \in D\}} \left[g(x) - g(x-)\right] \leq f(b) - f(a)$, existe $N \in \mathbb{N}$ tal que:

$\sum_{j=N+1}^{\infty} \left[g(x_j) - g(x_j-)\right] < \frac{\varepsilon}{4M}$.

Sean $z_1, z_2, \dots, z_N$ los elementos del conjunto $\{x_1, x_2, \dots, x_N\}$, ordenados del menor al mayor y definamos $x_0 = a$.

Para $j \in \{1, 2, \ldots, N\}$, tomemos $h \in (0, \text{mín}\{z_j - z_{j-1} : j \in \{1, 2, \ldots, N\}\})$ tal que:

$|f(y) - f(z_j)| < \frac{\varepsilon}{2[g(b)-g(a)]}$

para cualquier $j \in \{1, 2, \ldots, N\}$ y $y \in [z_j - h, z_j]$.

Sea $P_\varepsilon = \{x_0, z_1 - h, z_1, z_2 - h, z_2, \ldots, z_N - h, z_N, b\} = \{y_0, y_1, \ldots, y_m\}$, $P = \{u_0, u_1, \ldots, u_n\}$ un refinamiento de P_ε y, para cada $j \in \{1, 2, \ldots, N\}$, y_j el elemento de P más cercano a z_j por la izquierda. Como g crece únicamente mediante saltos, para cada $j \in \{1, 2, \ldots, n\}$, se tiene:

$g(u_j) - g(u_{j-1}) = \sum_{\{x \in D : x \in (a,b]\}} [g(x) - g(x-)]$.

Definamos:

$\mathcal{H} = \{[u_{j-1}, u_j] : j \in \{1, 2, \ldots, n\}\} - \{[y_j, z_j] : j \in \{1, 2, \ldots, n\}\}$.

Sea $S(P, f, g) = \sum_{j=1}^{n} f(\xi_j)[g(u_j) - g(u_{j-1})]$ una suma de Riemann-Stieltjes de f con respecto a g, correspondiente a la particion P.

Entonces:

$S(P, f, g) = \sum_{\{j \in \{1,2,\ldots,n\} : [u_{j-1}, u_j] \in \mathcal{H}\}} f(\xi_j)[g(u_j) - g(u_{j-1})] + \sum_{j=1}^{N} f(\gamma_j)[g(z_j) - g(y_j)]$.

Por lo tanto:

$\left| S(P, f, g) - \sum_{\{x \in D\}} f(x)[g(x) - g(x-)] \right|$

$\leq \left| \sum_{\{j \in \{1,2,\ldots,n\} : [u_{j-1}, u_j] \in \mathcal{H}\}} f(\xi_j)[g(u_j) - g(u_{j-1})] - \sum_{j=N+1}^{\infty} f(x_j)[g(x_j) - g(x_j-)] \right|$

$+ \left| \sum_{j=1}^{N} f(\gamma_j)[g(z_j) - g(y_j)] - \sum_{j=1}^{N} f(z_j)[g(z_j) - g(z_j-)] \right|$

$= \left| \sum_{\{j \in \{1,2,\ldots,n\} : [u_{j-1}, u_j] \in \mathcal{H}\}} f(\xi_j) \sum_{\{x \in D : x \in (u_{j-1}, u_j]\}} [g(x) - g(x-)] \right.$

$\left. + \sum_{j=N+1}^{\infty} f(x_j)[g(x_j) - g(x_j-)] \right|$

$+ \left| \sum_{j=1}^{N} f(\gamma_j) \sum_{\{x \in D : x \in (y_j, z_j)\}} [g(x) - g(x-)] + \sum_{j=1}^{N} f(\gamma_j)[g(z_j) - g(z_j-)] \right.$

$\left. - \sum_{j=1}^{N} f(z_j)[g(z_j) - g(z_j-)] \right|$

$= \left| \sum_{\{j \in \{1,2,\ldots,n\} : [u_{j-1}, u_j] \in \mathcal{H}\}} f(\xi_j) \sum_{\{x \in D : x \in (u_{j-1}, u_j]\}} [g(x) - g(x-)] \right.$

$\left. + \sum_{j=N+1}^{\infty} f(x_j)[g(x_j) - g(x_j-)] \right|$

$+\left|\sum_{j=1}^{N} f(\gamma_j) \sum_{\{x\in D: x\in(y_j,z_j)\}} [g(x)-g(x-)] + \sum_{j=1}^{N} \left[f(\gamma_j)-f(z_j)\right] [g(z_j)-g(z_j-)]\right|$

$\leq M \sum_{\{j\in\{1,2,\dots,n\}:[u_{j-1},u_j]\in\mathcal{H}\}} \sum_{\{x\in D: x\in(u_{j-1},u_j]\}} [g(x)-g(x-)] + M\sum_{j=N+1}^{\infty} [g(x_j)-g(x_j-)]$

$+M\sum_{j=1}^{N} \sum_{\{x\in D: x\in(y_j,z_j)\}} [g(x)-g(x-)] + \frac{\varepsilon}{2[g(b)-g(a)]}\sum_{j=1}^{N} [g(z_j)-g(z_j-)]$

$\leq 2M\sum_{j=N+1}^{\infty} [g(x_j)-g(x_j-)] + \frac{\varepsilon}{2[g(b)-g(a)]} [g(b)-g(a)]$

$< 2M\frac{\varepsilon}{4M} + \frac{\varepsilon}{2} = \varepsilon.$

■

De manera similar, se demuestra el siguiente resultado:

PROPOSICIÓN 4.7. *Si $g:[a,b]\to\mathbb{R}$ es una función no decreciente, continua por la izquierda y que crece únicamente mediante saltos y $f:[a,b]\to\mathbb{R}$ es una función acotada y continua por la derecha en los puntos donde g es discontinua, entonces f es integrable con respecto a g y se tiene $\int_a^b f dg = \sum_{\{x\in D\}} f(x)\,[g(x+)-g(x)]$, donde D es el conjunto de puntos en el intervalo $[a,b)$ donde g es discontinua.*

COROLARIO 4.5. *Si $f_1:[a,b]\to\mathbb{R}$ es una función no decreciente y continua por la derecha y $f_2:[a,b]\to\mathbb{R}$ es una función no decreciente y continua por la izquierda, entonces f_1 y f_2 son integrables una con respecto a la otra.*

Demostración

Por los corolarios 3.14 y 3.15, $f_1 = f_1^d + f_1^c$ y $f_2 = f_2^i + f_2^c$, donde f_1^c y f_2^c son funciones no decrecientes y continuas, f_1^d es una función no decreciente, continua por la derecha, que crece únicamente mediante saltos, y f_2^i es una función no decreciente, continua por la izquierda, que crece únicamente mediante saltos.

Por la proposición 4.6, $f_2^i + f_2^c$ es integrable con respecto a f_1^d y, por el corolario 4.4, es integrable con respecto a f_1^c, así que f_2 es integrable con respecto a f_1.

Finalmente, por el teorema 4.12, f_1 es integrable con respecto a f_2.

■

PROPOSICIÓN 4.8. *Si $g_1:[a,b]\to\mathbb{R}$ y $g_2:[a,b]\to\mathbb{R}$ son funciones de variación acotada, sin discontinuidades en común del mismo lado, entonces g_1 y g_2 son integrables una con respecto a la otra.*

Demostración

Por el teorema 3.4, $g_1 = f_1 - h_1$ y $g_2 = f_2 - h_2$, donde cada una de las parejas f_1, h_1 y f_1, h_1 está formada por funciones no decrecientes que no tienen discontinuidades en común del mismo lado.

Por la proposición 3.8, se tiene:

$f_1 = f_1^d + f_1^i + f_1^c$,

$h_1 = h_1^d + h_1^i + h_1^c$,

$f_2 = f_2^d + f_2^i + f_2^c$,

$h_2 = h_2^d + h_2^i + h_2^c$,

donde:

f_1^d, h_1^d, f_2^d y h_2^d son funciones no decrecientes, continuas por la derecha y que crecen únicamente mediante saltos.

f_1^i, h_1^i, f_2^i y h_2^i son funciones no decrecientes, continuas por la izquierda y quecrecen únicamente mediante saltos.

f_1^c, h_1^c, f_2^c y h_2^c son funciones no decrecientes y continuas.

Así que:

$g_1 = \left(f_1^d - h_1^d\right) + \left(f_1^i - h_1^i\right) + \left(f_1^c - h_1^c\right)$,

$g_2 = \left(f_2^d - h_2^d\right) + \left(f_2^i - h_2^i\right) + \left(f_2^c - h_2^c\right)$.

Por la proposición 4.6, f_1^i y h_1^i son integrables con respecto a f_2^d y h_2^d.

Por el corolario 4.7, f_1^d y h_1^d son integrables con respecto a f_2^i y h_2^i.

Por el corolario 4.4, f_1^d y h_1^d son integrables con respecto a f_2^c y h_2^c, y f_1^i y h_1^i son integrables con respecto a f_2^c y h_2^c.

Por el teorema 4.2, f_1^c y h_1^c son integrables con respecto a f_2^d, h_2^d, f_2^i, h_2^i, f_2^c y h_2^c.

Además:

Para cualquier $x \in \mathbb{R}$, se tiene:

$f_2^d(x) - f_2^d(x-) > 0 \Rightarrow h_2^d(x) - h_2^d(x-) = 0$,

$h_2^d(x) - h_2^d(x-) > 0 \Rightarrow f_2^d(x) - f_2^d(x-) = 0$,

$f_2^i(x+) - f_2^i(x) > 0 \Rightarrow h_2^i(x+) - h_2^i(x) = 0$,

$h_2^i(x+) - h_2^i(x) > 0 \Rightarrow f_2^i(x) - f_2^i(x-) = 0$.

Así que:

$f_2^d(x) - f_2^d(x-) > 0 \Rightarrow (f_2 - h_2)(x) - (f_2 - h_2)(x-) > 0$

$\Rightarrow g_2(x) - g_2(x-) > 0 \Rightarrow g_1(x) - g_1(x-) = 0 \Rightarrow f_1(x) - f_1(x-) = 0$ y $h_1(x) - h_1(x-) = 0$

$\Rightarrow \left(f_1^d - h_1^d\right)(x) - \left(f_1^d - h_1^d\right)(x-) = 0$

$h_2^d\,(x) - h_2^d\,(x-) > 0 \Rightarrow (f_2 - h_2)\,(x) - (f_2 - h_2)\,(x-) < 0$

$\Rightarrow g_2\,(x) - g_2\,(x-) < 0 \Rightarrow g_1\,(x) - g_1\,(x-) = 0 \Rightarrow f_1\,(x) - f_1\,(x-) = 0$ y $h_1\,(x) - h_1\,(x-) = 0$

$\Rightarrow \left(f_1^d - h_1^d\right)(x) - \left(f_1^d - h_1^d\right)(x-) = 0$

$f_2^i\,(x+) - f_2^i\,(x) > 0 \Rightarrow (f_2 - h_2)\,(x+) - (f_2 - h_2)\,(x) > 0$

$\Rightarrow g_2\,(x+) - g_2\,(x) > 0 \Rightarrow g_1\,(x+) - g_1\,(x) = 0 \Rightarrow f_1\,(x+) - f_1\,(x) = 0$ y $h_1\,(x+) - h_1\,(x) = 0$

$\Rightarrow (f_1^i - h_1^i)\,(x+) - (f_1^i - h_1^i)\,(x) = 0$

$h_2^i\,(x+) - h_2^i\,(x) > 0 \Rightarrow (f_2 - h_2)\,(x+) - (f_2 - h_2)\,(x) < 0$

$\Rightarrow g_2\,(x+) - g_2\,(x) < 0 \Rightarrow g_1\,(x+) - g_1\,(x) = 0 \Rightarrow f_1\,(x+) - f_1\,(x) = 0$ y $h_1\,(x+) - h_1\,(x) = 0$

$\Rightarrow (f_1^i - h_1^i)\,(x+) - (f_1^i - h_1^i)\,(x) = 0.$

Así que, por la proposición 4.6, $f_1^d - h_1^d$ es integrable con respecto a f_2^d y h_2^d, y, por la proposición 4.7, $f_1^i - h_1^i$ es integrable con respecto a f_2^i y h_2^i.

Por lo tanto, $\left(f_1^d - h_1^d\right) + (f_1^i - h_1^i) + (f_1^c - h_1^c)$ es integrable con respecto a f_2^d, h_2^d, f_2^i, h_2^i, f_2^c y h_2^c. Así que g_1 es integrable con respecto a g_2. Finalmente, por el teorema 4.12, g_2 es integrable con respecto a g_1.

■

CAPÍTULO 5

TEORÍA GENERAL DE LA MEDIDA

5.1. Introducción

La teoría de la medida y de la integración desarrollada por Lebesgue en su tesis doctoral causó un gran impacto, e inmediatamente después de su publicación, el mismo Lebesgue y otros investigadores comenzaron a uitilizarla en diferentes áreas, al mismo tiempo que se daban nuevos resultados, ampliando la teoría, y se generalizaban los conceptos definidos por Lebesgue.

> En 1905 Giuseppe Vitali ([**91**]) demostró que no es posible asignar una medida a todo conjunto acotado de números reales de tal manera que se satisfagan las condiciones que pide Lebesgue. Este resultado planteaba una disyuntiva, o bien se aceptan como razonables las condiciones de Lebesgue y entonces se restringe la familia de conjuntos a los cuales se les puede asignar una medida, o bien se buscan condiciones menos restrictivas para la medida de tal manera que pueda definirse para cualquier conjunto acotado de números reales. Se realizaron varios estudios al respecto, pero finalmente el medio matemático optó por mantener las condiciones de Lebesgue, aunque modificadas de tal manera que se pudiera eliminar la invarianza bajo traslaciones. Una cosa interesante del resultado de Vitali es que para demostrar la existencia de conjuntos que no son Lebesgue medibles, utiliza el axioma de elección. Sin la utilización de este axioma y sin agregar algún otro postulado a los axiomas de la matemática, no es posible demostrar la existencia de subconjuntos de números reales que no sean Lebesgue medibles, aunque tampoco es posible demostrar que todos los subconjuntos de números reales lo son ([**85**], [**71**]).

Después de que Lebesgue desarrolló su teoría de integración en $\mathbb{R}$, se extendió al caso de $\mathbb{R}^n$ sin mucha dificultad. Ya en su libro de 1904, Lebesgue había esbozado el caso de $\mathbb{R}^2$ y en 1910 desarrolló el caso general multidimensional.

En 1913, Johann Karl August Radon ([**76**]) mostró que se puede desarrollar una teoría general en la cual quedan incluidas la integral de Lebesgue en $\mathbb{R}^n$ y la integral de Stieltjes. Fue a partir de este trabajo que Fréchet logró desarrollar una formulación aún más general, de la cual surgiría la teoría general de la medida.

Radon introdujo el concepto de funcional aditiva sobre una familia de subconjuntos de $\mathbb{R}^n$. Una familia T, sobre la cual se definiría una funcional aditiva, debía tener las siguientes propiedades:

1. T contiene a todas las celdas acotadas de la forma $[a_1, b_1) \times [a_2, b_2) \times \cdots \times [a_n, b_n)$.

2. Si E_1 y E_2 pertenecen a T, entonces $E_1 \cap E_2 = E_1 \cap E_2$ y $E_1 - E_2$ también ´pertenecen a T.

3. Si $(E_n)_{n \in \mathbb{N}}$ es una sucesión de elementos de T, ajenos por parejas, entonces la unión de ellos también pertenece a T.

Obsérvese que si T es una familia con estas propiedades, entonces $\mathbb{R}^n$ pertenece a T. En efecto, $\mathbb{R}^n$ es la unión de todas las celdas de la forma $[m_1, m_1 + 1) \times [m_2, m_2 + 1) \times \cdots \times [m_n, m_n + 1)$, donde $m_1, m_2, \ldots, m_n$ son números enteros. Así que T constituye lo que ahora se conoce como σ-álgebra. Además, por la primera propiedad, T contiene a los conjuntos borelianos de $\mathbb{R}^n$.

Radon definió entonces una funcional aditiva como una función $f : T \to \mathbb{R}$ con la propiedad de que, si $(E_n)_{n \in \mathbb{N}}$ es una sucesión de elementos de T, ajenos por parejas, entonces:

$f\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} f\left(\bigcup_{n=1}^{\infty} E_n\right)$.

Aún no se trataba de la definición general de una medida ya que Radon se restringió a familias de subconjuntos de $\mathbb{R}^n$ que contienen a los conjuntos borelianos.

El concepto de funcional aditiva lo tomó Radon de un artículo de Lebesgue del año 1910 titulado *L'intégration des fonctions discontinues* ([**60**]), quien lo introdujo en el contexto del estudio del problema de determinar bajo que condiciones una función $F : [a, b] \to \mathbb{R}$ es una integral indefinida, es decir una función de la forma $x \to \int_a^x f(y)\, dy$, donde f es una función medible e integrable. Dentro de este estudio, Lebesgue tuvo la idea de considerar una integral indefinida como una función F que asigna a cada conjunto medible E la integral $\int_E f\,(P)\, dP$, donde f es una función medible e integrable y P representa un elemento de $\mathbb{R}^n$. Demostró entonces que, para una función así definida, si $(E_n)_{n \in \mathbb{N}}$ es una sucesión de conjuntos medibles, ajenos por parejas, entonces $F\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} F(E_n)$. A una función con esta propiedad, Lebesgue la denominó aditiva.

Siguiendo un procedimiento similar al de Lebesgue, Radon mostró que una función definida sobre el conjunto de celdas se puede extender, bajo determinadas condiciones, a una función aditiva.

Para el caso de $\mathbb{R}^n$, La función inicial puede estar definida por:

$f([a_1, b_1) \times [a_2, b_2) \times \cdots \times [a_n, b_n)) = \prod_{k=1}^{n} (b_k - a_k)$

Para el caso particular de $\mathbb{R}$, la función inicial puede estar definida por:

$f([a, b)) = G(b) - G(a)$,

donde G es una función de variación acotada continua por la izquierda.

Finalmente, también siguiendo un procedimiento similar al de Lebesgue, Radon desarrolló una teoría de integración para las funcionales aditivas.

Con base en el trabajo de Radon, Maurice René Fréchet extendió la teoría de la medida de Lebesgue a espacios abstractos en un artículo de 1915 titulado *Sur l'integrale d'une fonctionnelle ètendue a un ensemble abstrait* ([**37**]). Comenzó este artículo diciendo:

> *"M. J. Radon publicó recientemente (1913) una definición de la integral $\int F(P)\, dh(P)$ de una función $F(P)$ de un punto P del espacio de n dimensiones con respecto a una función de P, $h(P)$, de variación acotada. Esta definición resulta de una especie de fusiòn de la integral de Lebesgue y de la integral de Stieltjes. La definición de M. J. Radon se reduce a la de M. Lebesgue cuando h es una función lineal y a la de Stieltjes cuando F es una función continua. Por cierto, la integral de Radon puede tambièn escribirse:*
> $\int_E F(P)\, df(e)$,
> *donde $f(e)$ es una función aditiva del subconjunto variable e de E.*
>
> *Pero es bajo esta forma lo que me parece ser la gran ventaja de la definición de M. J. Radon, ventaja que no parece que él haya notado. M. J. Radon tenía como meta realizar un progreso en la teoría de funciones, unificando las definiciones de Stieltjes y de M. Lebesgue. Pero, de hecho, se nota que, con algunas ligeras modificaciones, la definición y las propiedades de la integral de M. Radon se extienden mucho màs allá del Cálculo integral clàsico, son casi inmediatamente aplicables al dominio infinitamente más vasto del Cálculo Funcional."*

Recordemos que Fréchet es uno de los precursores del Análisis Funcional. En 1906 presentó su tesis doctoral bajo el tìtulo *Sur quelques points du calcul fonctionnelle* ([**36**]), donde definió el concepto de métrica para un conjunto cualquiera. En su tesis doctoral y en trabajos posteriores desarrolló la teoría de los espacios métricos y en particular los espacios de funciones donde se puede definir una métrica, por ejemplo el conjunto de las funciones de cuadrado integrable con la métrica definida por $d(f, g) = \left(\int_a^b [f(x) - g(x)]^2\, dx\right)^{\frac{1}{2}}$.

En su artículo de 1915, Fréchet remarcaba que un conjunto es un conjunto abstracto cuando no conocemos la naturaleza de sus elementos o, lo que es lo mismo, cuando la naturaleza de sus elementos no interviene en los razonamientos que nos proponemos hacer sobre ese conjunto.

Definió una familia aditiva de conjuntos como una familia de conjuntos cerrada bajo diferencias y uniones finitas o infinito numerables.

Obsérvese que una familia aditiva es cerrada bajo intersecciones finitas o infinito numerables. En efecto, si $E_1, E_2, \ldots$ es una colección finita o infinita numerable de elementos de la familia y E es su unión, entonces:

$\bigcap_n E_n = E - \bigcup_n (E - E_n)$.

La única condición que hace falta para tener una σ-álgebra es que el conjunto total sea parte de la familia.

En seguida definió básicamente lo que después se llamaría una medida (incluyendo el caso de las medidas con signo):
Si $\Im$ es una familia aditiva y f es una función con valores reales, definida sobre $\Im$, se dice que f es aditiva si dada cualquier colección finita o infinita numerable, $E_1, E_2, \ldots$, de elementos de $\Im$, ajenos por parejas, entonces:
$f\left(\bigcup_n E_n\right) = \sum_n f(E_n)$.

Si $E_1, E_2, \ldots$ es una familia infinita numerable cualquiera, formada por elementos de $\Im$ ajenos por parejas, sean $F_1, F_2, \ldots$ los elementos de esa familia tales que f aplicada a cualquiera de ellos es un número real no negativo, y sean $G_1, G_2, \ldots$ los elementos de esa familia tales que f aplicada a cualquiera de ellos es un número real negativo. Se tiene entonces:

$f\left(\bigcup_n F_n\right) = \sum_n f(F_n)$.

$f\left(\bigcup_n G_n\right) = \sum_n f(G_n) = -\sum_n |f(G_n)|$.

Así que las series $\sum_n f(F_n)$ y $\sum_n |f(G_n)|$ son convergentes.

Por lo tanto, la serie $\sum_n |f(E_n)|$ es convergente; es decir, la serie $\sum_n f(E_n)$ es absolutamente convergente.

Finalmente, Fréchet definió la integral con respecto a una funciòn aditiva utilizando el método de Darboux, el cual consiste en definir la integral superior y la integral inferior.

En su artículo de 1923, al cual le siguió uno de 1924, Fréchet desarrolló aún más su teoría iniciada en 1915, quedando así ya establecido lo esencial de lo que posteriormente se llamaría la teoría de la medida y la teoría de integración con respecto a una medida.

Previamente, en el año 1914, Constantin Carathéodory, en un artículo titulado *Über das lineare Mass von Punktmengen - eine Verallgemeinerung des Längenbegriffs* (Sobre la medida lineal de los conjuntos de puntos- una generalización del concepto de longitud) ([**16**]), estableció un método para definir una medida a partir de una medida exterior en $\mathbb{R}^n$, el cual, prácticamente sin modificaciòn, se puede utilizar para definir una medida a partir de una medida exterior definida sobre los subconjuntos de un conjunto cualquiera.

> Una medida exterior m_e es una función definida sobre todos los subconjuntos de un conjunto F, la cual satisface las siguientes propiedades:
> $m_e(\emptyset) = 0$,
> Si $A \subset B$, entonces $m_e(A) \leq m_e(B)$,
> $m_e(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} m_e(A_n)$,
> Teniendo una medida exterior, se dice que un conjunto $E \subset F$ es medible si satisface la siguiente propiedad:
> $m_e(A) = m_e(A \cap E) + m_e(A \cap E^c)$
> para cualquier conjunto $A \subset F$.

El método de Carathéodory es el que quedó como estándar para definir una medida.

Podría decirse que el ciclo de investigación alrededor de los conceptos de medida y de integral con respecto a una medida, así como de sus propiedades básicas, se cerró con el trabajo de Otton Nikodym de 1930 ([**73**]), donde prácticamente expuso la formulación moderna de la teoría de la medida y de la teoría de la integración con respecto a una medida.

El camino seguido para llegar a esta formulación siguió fielmente la metodología de Lebesgue.

> Recordemos en síntesis como se dió el proceso: El problema central que trató Lebesgue fue el de la integración de funciones; así venía la historia desde que Cauchy formuló la definición de la integral para las funciones continuas; después vino el trabajo de Riemann, el cual marcó el camino posterior, siempre en la línea de resolver el problema de la integración de funciones. En las investigaciones alrededor de las condiciones para que una función sea Riemann-integrable surgió el concepto de contenido de un conjunto, lo cual fue ampliamente desarrollado por Jordan, mostrando que la teoría de integración se encuentra estrechamente vinculada con la teoría del contenido. Siguiendo este camino, Lebesgue encontró que, generalizando la teoría del contenido, para tener una familia más grande de subconjuntos de $\mathbb{R}$ a los cuales cuales se les puede asignar una medida, se obtiene también una generalización de la integral de Riemann, ampliando así la familia de funciones integrables.

Sin embargo, lo anterior es únicamente parte de la historia. No hay que olvidar que el objetivo de Lebesgue era resolver el problema de la integración de funciones. Para ello, se planteó el problema de encontrar una definición de la integral, la cual cumpliera determinadas condiciones (las 6 que mencionamos en el capítulo 1). Planteado de esta manera, algunos vieron el problema como uno de Análisis Funcional. Esta línea fue desarrollada por Frédéric Riesz ([**78**], [**79**], [**80**]), y, sobre todo, por Perci John Daniell en 4 artículos publicados entre 1918 y 1920 ([**21**], [**22**], [**23**], [**24**]), donde el problema de la integral lo planteó como un problema de extensión de una funcional lineal, por ejemplo la extensión de la funcional que asigna a cada función continua en un intervalo su integral (de Riemann) en ese intervalo. Con este enfoque logró definir el concepto de integral en espacios de dimensión infinita sin necesidad de construir una medida. En particular, su método fue utilizado por Norbert Wiener para construir un modelo matemático del movimiento browniano, aunque actualmente la manera usual de construir ese modelo es utilizando la teoría de la medida.

En este libro seguiremos el primer enfoque, así que a continuación expondremos la formulación moderna de la Teoría de la Medida.

5.2. Medidas sobre álgebras y σ-álgebras

Si analizamos la medida definida por Lebesgue, podemos ver que lo que obtuvo fue una función no negativa y σ-aditiva definida sobre una σ-álgebra de subconjuntos de $\mathbb{R}$. Esta idea queda recogida en la siguiente definición, con el agregado de que la medida del vacío es cero, propiedad que tiene la medida de Lebesgue y cualquier función μ no negativa y σ-aditiva definida sobre una σ-álgebra de subconjuntos de un conjunto dado para la cual exista un elemento E de la σ-álgebra tal que $\mu(E) < \infty$. En otras palabras, la condición $\mu(\emptyset) = 0$ es únicamente para excluir como medida a una función que asigne a todo elemento de la σ-álgebra el valor ∞.

Definición 5.1. *Sea $\mathbb{F}$ un conjunto y $\Im$ una σ-álgebra de subconjuntos de $\mathbb{F}$. Se dice que una función no negativa $\mu : \Im \mapsto \overline{\mathbb{R}}$ es una medida si μ es σ-aditiva y $\mu(\emptyset) = 0$.*

Definición 5.2. *Llamaremos espacio de medida a una terna $(\mathbb{F}, \Im, \mu)$ donde $\mathbb{F}$ es un conjunto, $\Im$ una σ-álgebra de subconjuntos de $\mathbb{F}$ y $\mu : \Im \mapsto \overline{\mathbb{R}}$ una medida.*

Definición 5.3. *Sea $(\mathbb{F}, \Im, \mu)$ un espacio de medida. Diremos que μ es finita si $\mu(\mathbb{F}) < \infty$. Diremos que μ es σ-finita si existe una colección infinita numerable de conjuntos $E_k \in \Im$ tales que $\mathbb{F} = \bigcup_{k=1}^{\infty} E_k$ y $\mu(E_k) < \infty$ para cualquier k.*

Si μ es σ-finita, los conjuntos $E_k \in \Im$ tales que $\mathbb{F} = \bigcup_{k=1}^{\infty} E_k$ y $\mu(E_k) < \infty$ para cualquier k, pueden escogerse de tal forma que sean ajenos por parejas. En efecto, los conjuntos $E_k' = E_k - \bigcup_{j=1}^{k-1} E_j$ son ajenos por parejas, $E_k' \in \Im$ y $\mu(E_k') < \infty$ para cualquier k, y $\mathbb{F} = \bigcup_{k=1}^{\infty} E_k'$. También pueden elegirse los conjuntos E_k de tal forma que la sucesión $(E_k)_{k\in\mathbb{N}}$ sea creciente. En efecto, si se tiene una sucesión de conjuntos $E_k \in \Im$ tales que $\mathbb{F} = \bigcup_{k=1}^{\infty} E_k$ y $\mu(E_k) < \infty$ para cualquier k, entonces, definiendo, para cualquier $k \in \mathbb{N}$, $E_k' = \bigcup_{j=1}^{k} E_j$, la sucesión $\left(E_k'\right)_{k\in\mathbb{N}}$ es creciente y tiene la misma propiedad.

En general, una medida se obtiene definiéndola primero para una familia de subconjuntos de un conjunto que no necesariamente forma una σ-álgebra; después se extiende a una familia más grande siguiendo el método de Lebesgue. Lo más común es buscar definirla sobre un álgebra de subconjuntos de un conjunto y después extenderla a la σ-álgebra generada por esa álgebra.

En el caso de la medida de Lebesgue, se puede ver que la propiedad básica que permite la extensión de la medida definida sobre los intervalos es el lema 2.1. En el caso general, la propiedad básica que permite extender una medida sobre un álgebra es la σ-subaditividad de esa medida.

DEFINICIÓN 5.4. *Sea $\mathbb{F}$ un conjunto y $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$. Se dice que una función no negativa $\mu : \mathcal{A} \mapsto \overline{\mathbb{R}}$ es σ-subaditiva, o que satisface la propiedad de la subaditividad numerable, si dada cualquier colección infinita $A_1, A_2, \ldots$ de elementos de $\mathcal{A}$ tales que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, entonces:*

$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

DEFINICIÓN 5.5. *Sea $\mathbb{F}$ un conjunto y $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$. Diremos que una función no negativa $\mu : \mathcal{A} \mapsto \overline{\mathbb{R}}$ es una quasi medida si es finitamente aditiva, σ-subaditiva y $\mu(\emptyset) = 0$.*

Como lo mencionamos antes, la condición de σ-subaditividad de una quasi medida definida sobre un álgebra $\mathcal{A}$ es la propiedad básica que se tiene que demostrar para poder extender esa medida a la σ-álgebra generada por $\mathcal{A}$. Para ese fin los siguientes dos resultados son útiles pues dan condiciones equivalentes de tal propiedad.

TEOREMA 5.1. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ una función no negativa y finitamente aditiva, entonces las siguientes propiedades son equivalentes:*

(i) *μ es σ-subaditiva.*
(ii) *μ es σ-aditiva.*
(iii) *Para cualquier colección infinita $A_1, A_2, \ldots$ de elementos de $\mathcal{A}$ tales que $A_1 \subset A_2 \subset \ldots$ y $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, se tiene $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightsquigarrow \infty} \mu(A_n)$.*

Demostración

$i \Longleftrightarrow ii$

Supongamos que μ es σ-subaditiva y sea $A_1, A_2, \ldots$ una colección numerable de elementos de $\mathcal{A}$, ajenos por parejas y tales que $\bigcup_j A_j \in \mathcal{A}$. Se tiene:

$\mu(\bigcup_j A_j) \geq \sum_{k=1}^{n} \mu(A_j)$ para cualquier $n \in \mathbb{N}$, así que $\mu(\bigcup_j A_j) \geq \sum_{k=1}^{\infty} \mu(A_j)$.

Por lo tanto, μ es σ-aditiva.

Supongamos ahora que μ es σ-aditiva y sea $A_1, A_2, \ldots$ una colección numerable de elementos de $\mathcal{A}$ tales que $\bigcup_j A_j \in \mathcal{A}$. Sea $B_n = A_n - \bigcup_{j=1}^{n-1} A_j$, entonces $\bigcup_j A_j = \bigcup_j B_j$, así que:

$\mu(\bigcup_j A_j) = \mu(\bigcup_j B_j) = \sum_{k=1}^{\infty} \mu(B_j) \leq \sum_{k=1}^{\infty} \mu(A_j)$.

$ii \Longrightarrow iii$

Supongamos que μ es σ-aditiva y sea $A_1, A_2, \ldots$ una colección de elementos de $\mathcal{A}$ tales que $A_1 \subset A_2 \subset \ldots$ y $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Si $\mu(A_k) = \infty$ para alguna $k \in \mathbb{N}$, el resultado es inmediato. Supongamos entonces que $\mu(A_k) < \infty$ para cualquier $k \in \mathbb{N}$. Se tiene entonces:

$\bigcup_{n=1}^{\infty} A_n = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \cdots$.

Así que:

$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{k=1}^{\infty} \mu(A_k - A_{k-1}) = \text{lím}_{n \rightsquigarrow \infty} \sum_{k=1}^{n} \mu(A_k - A_{k-1})$

$= \text{lím}_{n \rightsquigarrow \infty} \mu(A_n)$,

donde $A_0 = \emptyset$.

$iii \Longrightarrow i$

Sea $A_1, A_2, \ldots$ una colección numerable de elementos de $\mathcal{A}$ tales que $\bigcup_j A_j \in \mathcal{A}$. Para cada $n \in \mathbb{N}$, definamos $B_n = \bigcup_{j=1}^{n} A_j$, entonces $B_1 \subset B_2 \subset \ldots$ y $\bigcup_j A_j = \bigcup_j B_j$, así que:

$\mu(\bigcup_j A_j) = \mu(\bigcup_j B_j) = \text{lím}_{n \to \infty} \mu(B_n) \leq \text{lím}_{n \rightsquigarrow \infty} \sum_{j=1}^{n} \mu(A_j) = \sum_{k=1}^{\infty} \mu(A_j)$.

■

En la siguiente proposición, la función μ es finita para cualquier elemento del álgebra $\mathcal{A}$; de ahí que se tengan algunas propiedades adicionales a las de la proposición anterior.

TEOREMA 5.2. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y $\mu : \mathcal{A} \to \mathbb{R}$ una función no negativa y finitamente aditiva, entonces las siguientes propiedades son equivalentes:*

(i) *μ es σ-subaditiva.*
(ii) *μ es σ-aditiva.*
(iii) *Para cualquier sucesión creciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\mathcal{A}$ tales que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, se tiene $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \text{lím}_{n \rightsquigarrow \infty} \mu(A_n)$.*
(iv) *Para cualquier sucesión decreciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\mathcal{A}$ tales que $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$, se tiene $\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \text{lím}_{n \rightsquigarrow \infty} \mu(A_n)$.*
(v) *Para cualquier sucesión decreciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\mathcal{A}$ tales que $\bigcap_{n=1}^{\infty} A_n = \emptyset$, se tiene $\text{lím}_{n\to\infty} \mu(A_n) = 0$.*

Demostración

La equivalencia de i, ii y iii ya se demostró para cualquier función $\mu : \mathcal{A} \to \mathbb{R}^+ \cup \{\infty\}$ función finitamente aditiva.

$iii \Longrightarrow iv$ es inmediato tomando complementos.

$iv \Longrightarrow v$ es inmediato pues v es un caso particular de iv.

$v \Longrightarrow ii$

Sea $A_1, A_2, \ldots$ una colección numerable de elementos de $\mathcal{A}$, ajenos por parejas y tales que $\bigcup_j A_j \in \mathcal{A}$. Sea $B_n = \bigcup_{j=1}^{\infty} A_j - \bigcup_{j=1}^{n} A_j$, entonces $B_1 \supset B_2 \supset \ldots$ y $\bigcap_{n=1}^{\infty} B_n = \emptyset$.

Por lo tanto, $\lim_{n\rightsquigarrow\infty} \mu(B_n) = 0$, así que:

$\mu(\bigcup_{j=1}^{\infty} A_j) = \lim_{n\rightsquigarrow\infty} \mu(\bigcup_{j=1}^{n} A_j) = \lim_{n\rightsquigarrow\infty} \sum_{k=1}^{n} \mu(A_j) = \sum_{k=1}^{\infty} \mu(A_j)$.

■

Teorema 5.3. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ una quasi medida. Entonces, para cualquier sucesión decreciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\mathcal{A}$, tales que $\mu(A_N) < \infty$ para alguna $N \in \mathbb{N}$ y $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$, se tiene:*

$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\rightsquigarrow\infty} \mu(A_n)$.

Demostración

Sea $(A_n)_{n\in\mathbb{N}}$ sucesión decreciente de elementos de $\mathcal{A}$ tales que $\mu(A_N) < \infty$ para alguna $N \in \mathbb{N}$ y $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$.

Para cada $k \in \{N+1, N+2, \ldots\}$, definamos $B_k = A_N - A_k$. Entonces la sucesión $(B_{N+n})_{n\in\mathbb{N}}$ es creciente y:

$\bigcup_{n=N+1}^{\infty} B_n = A_N - \bigcap_{n=N+1}^{\infty} A_n \in \mathcal{A}$.

Así que:

$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \mu\left(\bigcap_{n=N+1}^{\infty} A_n\right) = \mu(A_N) - \mu\left(\bigcup_{n=N+1}^{\infty} B_n\right)$

$= \mu(A_N) - \lim_{n\rightsquigarrow\infty} \mu(A_N - A_{N+n}) = \mu(A_N) = \lim_{n\to\infty} \mu(A_{N+n}) = \lim_{n\rightsquigarrow\infty} \mu(A_n)$.

■

Como corolario, se tiene el siguiente resultado:

Teorema 5.4. *Sea $(\mathbb{F}, \Im, \mu)$ un espacio de medida. Entonces:*

(i) *μ es σ-subaditiva.*
(ii) *Para cualquier sucesión creciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\Im$, se tiene $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\rightsquigarrow\infty} \mu(A_n)$.*
(iii) *Para cualquier sucesión decreciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\Im$ tales que $\mu(A_N) < \infty$ para alguna $N \in \mathbb{N}$, se tiene $\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\rightsquigarrow\infty} \mu(A_n)$.*

Las propiedades ii y iii del teorema anterior dan la idea de que una medida es continua en un cierto sentido, aunque rigurosamente, para poder hablar de la continuidad de una medida se requiere que en el dominio donde está definida, en este caso una σ-álgebra de subconjuntos de un conjunto, se tenga definida una topología que nos permita hablar de vecindades y límites en ese dominio. No la tenemos definida; sin embargo, podemos definir lo que entendemos por el límite de una sucesión de conjuntos, cuando existe, basándonos en conceptos similares a los de límite superior e inferior de una sucesión de números reales.

DEFINICIÓN 5.6. *Si* $\mathbb{E}$ *es un conjunto y* $(A_n)_{n\in\mathbb{N}}$ *una sucesión de subconjuntos de* $\mathbb{E}$*, definimos:*

$\liminf_{n\rightsquigarrow\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m,$

$\limsup_{n\rightsquigarrow\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m,$

Si $\limsup_{n\rightsquigarrow\infty} A_n = \liminf_{n\to\infty} A_n$*, se dice que la sucesión* $(A_n)_{n\in\mathbb{N}}$ *converge y al valor común de* $\liminf_{n\rightsquigarrow\infty} A_n$ *y* $\limsup_{n\rightsquigarrow\infty} A_n$ *se le denota por* $\lim_{n\to\infty} A_n$*.*

Obsérvese que el límite inferior de una sucesión $(A_n)_{n\in\mathbb{N}}$ de subconjuntos de un conjunto $\mathbb{E}$ es el conjunto formado por todos los elementos $x\in\mathbb{E}$ tales que existe $N\in\mathbb{N}$ tal que $x\in A_n$ para cualquier $n\geq N$.

Por su parte, el límite superior de una sucesión $(A_n)_{n\in\mathbb{N}}$ de subconjuntos de un conjunto $\mathbb{E}$ es el conjunto formado por todos los elementos $x\in\mathbb{E}$ tales que $x\in A_n$ para una infinidad de números naturales n.

Obsérvese también que si la sucesión $(A_n)_{n\in\mathbb{N}}$ es creciente, entonces $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$; mientras que si $(A_n)_{n\in\mathbb{N}}$ es decreciente, entonces $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

TEOREMA 5.5. *Sea* $(\mathbb{F},\Im,\mu)$ *un espacio de medida. Supongamos que* μ *es finita y sea* $(A_n)_{n\in\mathbb{N}}$ *una sucesión de elementos de* $\Im$ *tal que* $\limsup_{n\rightsquigarrow\infty} A_n = \liminf_{n\rightsquigarrow\infty} A_n$*, entonces la sucesión* $(\mu(A_n))_{n\in\mathbb{N}}$ *es convergente y:*

$\mu(\lim_{n\rightsquigarrow\infty} A_n) = \lim_{n\rightsquigarrow\infty}\mu(A_n).$

Demostración

Definamos $A=\lim A_n$.

Para cada $n\in\mathbb{N}$ definamos $B_n=\bigcup_{m=n}^{\infty} A_m$ y $C_n=\bigcap_{m=n}^{\infty} A_m$. Entonces, la sucesión $(B_n)_{n\in\mathbb{N}}$ es decreciente y la sucesión $(C_n)_{n\in\mathbb{N}}$ es creciente. Además:

$\mu(\lim_{n\rightsquigarrow\infty} A_n) = \mu(\liminf_{n\rightsquigarrow\infty} A_n) = \mu(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m) = \mu(\bigcup_{n=1}^{\infty} C_n) = \lim_{n\rightsquigarrow\infty}\mu(C_n),$

$\mu(\lim_{n\rightsquigarrow\infty} A_n) = \mu(\limsup_{n\rightsquigarrow\infty} A_n) = \mu(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m) = \mu(\bigcap_{n=1}^{\infty} B_n) = \lim_{n\rightsquigarrow\infty}\mu(B_n).$

Por otra parte, para cualquier $n\in\mathbb{N}$, $B_n\supset A_n$ y $C_n\subset A_n$, así que:

$\mu(C_n)\leq\mu(A_n)\leq\mu(B_n)$ para cualquier $n\in\mathbb{N}$.

Por lo tanto:

$\mu\left(\text{lím}_{n\rightsquigarrow\infty} A_n\right) = \text{lím}_{n\rightsquigarrow\infty} \mu\left(C_n\right) \leq \text{lím inf}_{n\rightsquigarrow\infty} \mu\left(A_n\right) \leq \text{lím sup}_{n\rightsquigarrow\infty} \mu\left(A_n\right)$

$\leq \text{lím}_{n\rightsquigarrow\infty} \mu\left(B_n\right) = \mu\left(\text{lím}_{n\rightsquigarrow\infty} A_n\right)$.

Así que, $\text{lím inf}_{n\rightsquigarrow\infty} \mu\left(A_n\right) = \text{lím sup}_{n\rightsquigarrow\infty} \mu\left(A_n\right) = \mu\left(\text{lím}_{n\rightsquigarrow\infty} A_n\right)$.

Por lo tanto, la sucesión $\left(\mu\left(A_n\right)\right)_{n\in\mathbb{N}}$ es convergente y:

$\text{lím}_{n\rightsquigarrow\infty} \mu\left(A_n\right) = \mu\left(\text{lím}_{n\rightsquigarrow\infty} A_n\right)$.

■

5.3. Construcción de medidas

En esta sección $\mathbb{F}$ será un conjunto cualquiera fijo, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y μ_0 una quasi medida sobre $\mathcal{A}$. Todos los conjuntos con los que trataremos serán subconjuntos de $\mathbb{F}$.

Vamos a ver cómo, siguiendo el método de Lebesgue, se puede extender una quasi medida, definida sobre un álgebra $\mathcal{A}$ de subconjuntos de un conjunto $\mathbb{F}$, a la σ-álgebra, $\sigma(\mathcal{A})$, generada por el álgebra. Primero definiremos la medida exterior de cualquier subconjunto de $\mathbb{F}$; después definiremos la medibilidad de un conjunto utilizando el criterio de Carathéodory. Una vez hecho esto, mostraremos que la familia de conjuntos medibles forma una σ-álgebra, la cual contiene a los elementos de $\mathcal{A}$ y a los conjuntos de medida exterior cero. La medida de un conjunto medible la definiremos como su medida exterior y mostraremos que la medida así definida, restringida a $\mathcal{A}$, coincide con μ_0.

DEFINICIÓN 5.7. *Diremos que una colección finita o infinita numerable $A_1, A_2, \ldots$ de elementos de $\mathcal{A}$ es una cubierta del conjunto A si $A \subset \bigcup_n A_n$.*

DEFINICIÓN 5.8. *Se define la medida exterior, $\mu_e(A)$, de un conjunto A, mediante la relación*

$\mu_e(A) = \text{ínf}\left\{\sum_j \mu_0(A_j) : A_1, A_2, \ldots \text{ es cubierta de } A\right\}$.

PROPOSICIÓN 5.1. *Si A y B son dos conjuntos tales que $A \subset B$ entonces $\mu_e(A) \leq \mu_e(B)$.*

Gracias a la σ-subaditividad de μ_0 se tiene el siguiente resultado:

PROPOSICIÓN 5.2. *Si $A \in \mathcal{A}$ entonces $\mu_e(A) = \mu_0(A)$.*

Demostración

Sea $A \in \mathcal{A}$ y $A_1, A_2, \ldots$ una cubierta de A, entonces $A_n \cap A \in \mathcal{A}$ para cualquier $n \in \mathbb{N}$ y $A = \bigcup_n (A_n \cap A)$; así que, como μ_0 es σ-subaditiva:

$\mu_0(A) \leq \sum_n \mu_0(A_n \cap A) \leq \sum_n \mu_0(A_n)$

Por lo tanto, como esto ocurre para cualquier cubierta de A, $\mu_0(A) \leq \mu_e(A)$.

Por otra parte, como A es una cubierta de él mismo, se tiene $\mu_e(A) \leq \mu_0(A)$.

■

PROPOSICIÓN 5.3. *Si $(A_n)_{n\in\mathbb{N}}$ es una sucesión de conjuntos, entonces:*

$\mu_e\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu_e(A_n)$.

Demostración

Si $\mu_e(A_n) = \infty$ para alguna n el resultado es trivial.

Supongamos entonces que $\mu_e(A_n) < \infty$ para toda n. Dada $\varepsilon > 0$, para cada conjunto A_n sea $A_{n1}, A_{n2}, \ldots$ una cubierta de A_n tal que $\sum_m \mu_0(A_{nm}) < \mu_e(A_n) + \frac{\varepsilon}{2^n}$. La familia de conjuntos A_{nm} forman una cubierta de $\bigcup_n A_n$, así que:

$\mu_e\left(\bigcup_n A_n\right) \leq \sum_n \sum_m \mu_0(A_{nm}) \leq \sum_n \left[\mu_e(A_n) + \frac{\varepsilon}{2^n}\right] \leq \sum_n \mu_e(A_n) + \varepsilon;$

es decir, $\mu_e\left(\bigcup_n A_n\right) \leq \sum_n \mu_e(A_n) + \varepsilon$ para cualquier $\varepsilon > 0$. Por lo tanto:

$\mu_e\left(\bigcup_n A_n\right) \leq \sum_n \mu_e(A_n)$.

■

DEFINICIÓN 5.9. *Diremos que un conjunto E es medible si $\mu_e(A) = \mu_e(A \cap E) + \mu_e(A \cap E^c)$ para cualquier conjunto A. Además, en este caso, se define la medida de E, $\mu(E)$, como la medida exterior de E.*

Obsérvese que, por la σ-subaditividad de la medida exterior, se tiene:

$$\mu_e(A) \leq \mu_e(A \cap E) + \mu_e(A \cap E^c)$$

para cualquier par de conjuntos E y A, de manera que para demostrar la medibilidad de un conjunto E únicamente es necesario probar la otra desigualdad.

PROPOSICIÓN 5.4. *La familia de conjuntos medibles forma un álgebra de subconjuntos de $\mathbb{F}$.*

Demostración

Que el conjunto $\mathbb{F}$ es medible, así como que el complemento de un conjunto medible es medible, son resultados obvios.

Sean E_1 y E_2 dos conjuntos medibles y A cualquier conjunto. Se tiene entonces:

$\mu_e\left(A \cap (E_1 \cup E_2)\right) + \mu_e\left(A \cap (E_1 \cup E_2)^c\right)$

$= \mu_e\left((A \cap E_1) \cup (A \cap E_1^c \cap E_2)\right) + \mu_e\left(A \cap E_1^c \cap E_2^c\right)$

$\leq \mu_e(A \cap E_1) + \mu_e(A \cap E_1^c \cap E_2) + \mu_e\left(A \cap E_1^c \cap E_2^c\right)$

$= \mu_e(A \cap E_1) + \mu_e(A \cap E_1^c) = \mu_e(A)$.

Así que, $E_1 \cup E_2$ es medible.

■

PROPOSICIÓN 5.5. *La función que asigna a cada conjunto medible E su medida, $\mu(E)$, es una función finitamente aditiva.*

Demostración

Sean E_1 y E_2 dos conjuntos medibles ajenos, entonces, como $E_1 \cup E_2$ es medible, se tiene:

$$\mu\left(E_1 \cup E_2)\right) = \mu\left((E_1 \cup E_2)) \cap E_1\right) + \mu\left((E_1 \cup E_2) \cap E_1^c\right)$$

$$= \mu\left(E_1\right) + \mu\left(E_2\right).$$

■

PROPOSICIÓN 5.6. *La familia de conjuntos medibles forma una σ-álgebra de subconjuntos de $\mathbb{F}$.*

Demostración

Sea $E_1, E_2, \ldots$ una colección infinita numerable de conjuntos medibles ajenos por parejas y A cualquier subconjunto de $\mathbb{F}$.

Demostremos que $\mu_e\left(A \cap (\bigcup_{j=1}^{n} E_j)\right) = \sum_{j=1}^{n} \mu_e(A \cap E_j)$ para cualquier $n \in \mathbb{N}$.

Para $n = 1$ la igualdad es obvia.

Supongamos ahora que la igualdad es válida para $n = k$, entonces, como E_{k+1} es medible, se tiene:

$$\mu_e\left(A \cap (\bigcup_{j=1}^{k+1} E_j)\right) = \mu_e\left(A \cap (\bigcup_{j=1}^{k+1} E_j) \cap E_{k+1}\right) + \mu_e\left(A \cap (\bigcup_{j=1}^{k+1} E_j) \cap E_{k+1}^c\right)$$

$$= \mu_e\left(A \cap E_{k+1}\right) + \mu_e\left(A \cap (\bigcup_{j=1}^{k} E_j)\right) = \mu_e\left(A \cap E_{k+1}\right) + \sum_{j=1}^{k} \mu_e(A \cap E_j)$$

$$= \sum_{j=1}^{k+1} \mu_e(A \cap E_j).$$

Por lo tanto, la igualdad es válida para $n = k + 1$, así que, por el principio de inducción, lo es para cualquier $n \in \mathbb{N}$.

Ahora bien, como la familia de conjuntos medibles forma un álgebra de subconjuntos de $\mathbb{F}$, para cada $n \in \mathbb{N}$ el conjunto $\bigcup_{j=1}^{n} E_j$ es medible, así que:

$$\mu_e(A) = \mu_e\left(A \cap (\bigcup_{j=1}^{n} E_j)\right) + \mu_e\left(A \cap (\bigcup_{j=1}^{n} E_j)^c\right)$$

$$= \sum_{j=1}^{n} \mu_e(A \cap E_j) + \mu_e\left(A \cap (\bigcup_{j=1}^{n} E_j)^c\right)$$

$\geq \sum_{j=1}^{n} \mu_e(A \cap E_j) + \mu_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)^c\right)$.

Tomando límite cuando $n \rightsquigarrow \infty$ y utilizando la σ-subaditividad de la medida exterior,se obtiene entonces:

$\mu_e(A) \geq \sum_{j=1}^{\infty} \mu_e(A \cap E_j) + \mu_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)^c\right)$

$\geq \mu_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)\right) + \mu_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)^c\right)$.

Por lo tanto, $\bigcup_{j=1}^{\infty} E_j$ es medible.

∎

Proposición 5.7. *La función que asigna a cada conjunto medible E su medida, $\mu(E)$, es una función σ-aditiva.*

Demostración

Sea $E_1, E_2, \ldots$ una colección infinita numerable de conjuntos medibles ajenos por parejas. Por la σ-subaditividad de la medida exterior, se tiene $\mu(\bigcup_{j=1}^{\infty} E_j) \leq \sum_{j=1}^{\infty} \mu(E_j)$. Por otra parte, por la aditividad finita de la función que asigna a cada conjunto medible su medida, se tiene, para cualquier $n \in \mathbb{N}$:

$\mu(\bigcup_{j=1}^{\infty} E_j) \geq \mu(\bigcup_{j=1}^{n} E_j) = \sum_{j=1}^{n} \mu(E_j)$.

Así que tomando límite cuando $n \rightsquigarrow \infty$, se Aiene:

$\mu(\bigcup_{j=1}^{\infty} E_j) \geq \sum_{j=1}^{\infty} \mu(E_j)$.

∎

Si denotamos por $\mathcal{M}$ a la familia de los conjuntos medibles, sabemos ya que $\mathcal{M}$ forma una σ-álgebra de subconjuntos de $\mathbb{F}$. Además, la función $\mu : \mathcal{M} \mapsto \overline{R}$ es no negativa, σ-aditiva y $\mu(\emptyset) = 0$. Así que μ es un a medida definida sobre una σ-álgegra de subconjuntos de $\mathbb{F}$. Lo que resta probar es que μ es una extensión de μ_0. Vamos a demostrar que efectivamente esto es así y veremos que $\mathcal{M}$ es más grande que la σ-álgebra generada por $\mathcal{A}$. Esto es análogo a lo que ocurre con la medida de Lebesgue, la cual está definida no únicamente sobre los subconjuntos borelianos de $\mathbb{R}$, sino también sobre todos los conjuntos de medida cero.

Proposición 5.8. *Todo conjunto de medida exterior cero es medible.*

Demostración

Sea E un conjunto de medida exterior cero y A cualquier conjunto, entonces $A \cap E$ tiene medida exterior cero, así que:

$\mu_e(A) \geq \mu_e(A \cap E^c) = \mu_e(A \cap E^c) + \mu_e(A \cap E)$.

∎

Proposición 5.9. *Todo elemento de $\mathcal{A}$ es medible.*

Demostración

Sea $E \in \mathcal{A}$, A cualquier conjunto y $A_1, A_2, \dots$ una cubierta de A, entonces, para cada A_n, los conjuntos $A_n \cap E$ y $A_n \cap E^c$ pertenecen a $\mathcal{A}$ y se tiene:

$\mu_e(A \cap E) \leq \mu_e\left(\left(\bigcup_n A_n\right) \cap E\right) = \mu_e\left(\bigcup_n (A_n \cap E)\right)$

$\leq \sum_n \mu_e(A_n \cap E) = \sum_n \mu_0(A_n \cap E)$,

$\mu_e(A \cap E^c) \leq \mu_e\left(\left(\bigcup_n A_n\right) \cap E^c\right) = \mu_e\left(\bigcup_n (A_n \cap E^c)\right)$

$\leq \sum_n \mu_e(A_n \cap E^c) = \sum_n \mu_0(A_n \cap E^c)$.

Así que:

$\mu_e(A \cap E) + \mu_e(A \cap E^c) \leq \sum_n \mu_0(A_n \cap E) + \sum_n \mu_0(A_n \cap E^c) = \sum_n \mu_0(A_n)$.

Finalmente, como lo anterior es válido para cualquier cubierta de A, se puede concluir que:

$\mu_e(A \cap E) + \mu_e(A \cap E^c) \leq \mu_e(A)$.

■

Proposición 5.10. *Todo elemento de $\sigma(\mathcal{A})$ es medible.*

Demostración

El resultado es inmediato pues la familia de conjuntos medibles forma una σ-álgebra que contiene a los elementos de $\mathcal{A}$.

■

Los resultados anteriores pueden condensarse en el siguiente:

Teorema 5.6 (Teorema de extensión de Carathéodory). *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y $\mu_0 : \mathcal{A} \mapsto \mathbb{R}^+ \cup \{\infty\}$ una quasi medida. Entonces existe una medida $\mu : \Im \mapsto \mathbb{R}^+ \cup \{\infty\}$ tal que $\mu(A) = \mu_0(A)$ para cualquier $A \in \mathcal{A}$, donde $\Im$ es una σ-álgebra que contiene a $\sigma(\mathcal{A})$ y a los conjuntos de medida exterior cero.*

Definición 5.10. *Si $\mathbb{F}$ es un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y $\mu_0 : \mathcal{A} \mapsto \mathbb{R} \cup \{\infty\}$ una quasi medida, a la medida $\mu : \Im \mapsto \mathbb{R}^+ \cup \{\infty\}$ del teorema anterior la llamaremos la medida generada por la quasi medida μ_0.*

Ahora veremos que la familia de conjuntos medibles $\Im$, si bien es más grande que la σ-álgebra generada por $\mathcal{A}$, en general no es más grande que la σ-álgebra generada por $\mathcal{A}$ y los subconjuntos de $\mathbb{F}$ que tienen medida exterior cero.

Obsérvese que si un conjunto $B \subset \mathbb{F}$ tiene medida exterior cero, entonces, dada cualquier $\varepsilon > 0$ existe una colección infinita numerable, $A_1, A_2 \ldots$, de elementos de $\mathcal{A}$ tales que $B \subset \bigcup_{k=1}^{\infty} A_k$ y $\sum_{k=1}^{\infty} \mu_0(A_k) < \varepsilon$.

PROPOSICIÓN 5.11. *Dado cualquier conjunto medible E, de medida finita, existe $B \in \sigma(\mathcal{A})$ y un conjunto C, de medida exterior cero, tales que $E = B \cup C$ y $B \cap C = \emptyset$.*

Demostración

Sea E un conjunto medible de medida finita y, dada $\varepsilon > 0$, sea $A_1, A_2, \ldots$ una cubierta de E tal que:

$\sum_j \mu(A_j) < \mu(E) + \varepsilon$.

$A_\varepsilon = \bigcup_j A_j$ es entonces un elemento $\sigma(\mathcal{A})$ tal que:

$\mu(A_\varepsilon - E) = \mu(A_\varepsilon) - \mu(E) \leq \sum_j \mu(A_j) - \mu(E) < \varepsilon$.

Es decir, dada $\varepsilon > 0$ existe $A_\varepsilon \in \sigma(\mathcal{A})$ tal que $A_\varepsilon \supset E$ y $\mu(A_\varepsilon - E) < \varepsilon$.

Sea entonces $(B_n)_{n\in\mathbb{N}}$ una sucesión decreciente de elementos de $\sigma(\mathcal{A})$, que contengan a E, tales que $\mu(A_n - E) < \frac{1}{n}$ para cualquier $n \in \mathbb{N}$. Se tiene entonces $\mu\left(\bigcap_{j=1}^{\infty} A_j - E\right) \leq \mu(A_n - E) < \frac{1}{n}$ para cualquier $n \in \mathbb{N}$, así que $\mu\left(\bigcap_{j=1}^{\infty} A_j - E\right) = 0$.

Por lo tanto, existe $A \in \sigma(\mathcal{A})$ tal que $A \supset E$ y $\mu(A - E) = 0$.

Sea entonces $D \in \sigma(\mathcal{A})$ tal que $D \supset E^c$ y $\mu(D - E^c) = 0$. Entonces $B = D^c$ es un elemento de $\sigma(\mathcal{A})$ tal que $B \subset E$ y $\mu(E - B) = \mu(E - B^c) = \mu(E \cap B) = \mu(B - E^c) = 0$, de manera que $B \in \sigma(\mathcal{A})$, $F = E - B$ es un conjunto de medida exterior cero y se tiene $E = B \cup F$.

Finalmente, definamos $C = F - B$, entonces C tiene medida exterior cero y se tiene $E = B \cup C$ y $B \cap C = \emptyset$.

∎

COROLARIO 5.1. *Si μ es σ-finita, entonces, dado cualquier conjunto medible E existe $B \in \sigma(\mathcal{A})$ y un conjunto C, de medida exterior cero, tales que $E = B \cup C$ y $B \cap C = \emptyset$.*

Demostración

Sea $(F_k)_{k\in\mathbb{N}}$ una sucesión creciente de conjuntos medibles tales que $\mathbb{F} = \bigcup_{k=1}^{\infty} F_k$ y $\mu(F_k) < \infty$ para cualquier $k \in \mathbb{N}$.

Para cada $k \in \mathbb{N}$, definamos $E_k = E \cap F_k$.

Sea $B_k \in \sigma(\mathcal{A})$ y C_k un conjunto de medida exterior cero tales que $E_k = B_k \cup C_k$ y $B_k \cap C_k = \emptyset$, entonces tomando $B = \bigcup_{k=1}^{\infty} B_k$ y $D = \bigcup_{k=1}^{\infty} C_k$, se tiene que $B \in \sigma(\mathcal{A})$, D tiene medida exterior cero y $E = B \cup D$.

Finalmente, definamos $C = D - B$, entonces C tiene medida exterior cero y se tiene $E = B \cup C$ y $B \cap C = \emptyset$.

■

Corolario 5.2. *Si μ es σ-finita, la σ-álgebra de los conjuntos medibles es la σ-álgebra generada por $\mathcal{A}$ y los conjuntos de medida exterior cero.*

Proposición 5.12. *Todo conjunto de medida exterior cero está contenido en un conjunto $B \in \sigma(\mathcal{A})$ de medida exterior cero.*

Demostración

Sea A un conjunto de medida exterior cero. Para cada $n \in \mathbb{N}$, sea $\{A_k^n\}$ una colección infinita numerable de elementos de $\mathcal{A}$ tales que $A \subset \bigcup_{k=1}^{\infty} A_k^n$ y $\sum_{k=1}^{\infty} \mu_0(A_k^n) < \frac{1}{n}$. Definamos $B_n = \bigcup_{k=1}^{\infty} A_k^n$ y $B = \bigcap_{n=1}^{\infty} B_n$. Entonces, $B \in \sigma(\mathcal{A})$, tiene medida exterior cero y $A \subset B$.

■

Definición 5.11. *Sea $\mathbb{F}$ un conjunto, $\Im$ una σ-álgebra de subconjuntos de $\mathbb{F}$ y μ una medida sobre $\Im$. Diremos que $\Im$ es completa con respecto a μ si contiene a todos los subconjuntos de los conjuntos de medida μ igual a cero.*

Si una σ-álgebra $\Im_0$ no es completa con respecto a una medida μ, se puede completar. En efecto, sea H la familia de conjuntos $B \in \Im_0$ tales que $\mu(B) = 0$, entonces la familia $\Im$ de conjuntos de la forma $A \cup E$, donde $A \in \Im_0$ y E es un subconjunto de un conjunto $B \in H$, forma una σ-álgebra de subconjuntos de F.

La prueba de que $\mathbb{F} \in \Im$ y que $\Im$ es cerrada bajo uniones numerables es inmediata. Para probar que $\Im$ es cerrada bajo complementos, sea $C = A \cup E \in \Im$, donde $A \in \Im_0$ y E es un subconjunto de un conjunto $B \in H$. Entonces:

$$C^c = (A \cup E)^c = A^c \cap E^c = A^c \cap [E^c \cap (B \cup B^c)]$$

$$= A^c \cap [(E^c \cap B) \cup B^c] = (A^c \cap B^c) \cup (A^c \cap E^c \cap B).$$

Así que $C^c \in \Im$.

Obsérvese que $\Im$ es la σ-álgebra generada por $\Im_0$ y los subconjuntos de conjuntos $B \in \Im_0$ de medida cero.

Extendamos μ a $\Im$ definiendo $\mu(A \cup E) = \mu(A)$ para cualesquiera $A \in \Im_0$ y E un subconjunto de un conjunto $B \in H$.

Sea $C \in \Im$ tal que $\mu(C) = 0$, entonces $C = A \cup E$, en donde $A \in \Im_0$ y $E \subset B$, con $B \in H$. Así que:

Como $\mu(A) = \mu(A \cup E) = \mu(C) = 0$, entonces $A \in H$.

Sea ahora $D \subset C$, entonces $D \subset A \cup B$, así que $D \in \Im$. Es decir, $\Im$ es completa con respecto a μ.

Lo anterior también muestra que si $C \in \Im$ y $\mu(C) = 0$, entonces $C \subset A \cup B$, con $A, B \in H$. Por lo tanto, todo conjunto $C \in \Im$ de medida cero está contenido en un conjunto $G \in \Im_0$ de medida cero.

Retomando el enunciado del teorema 5.6, la medida μ restringida a $\sigma(\mathcal{A})$ sigue siendo una medida. La proposición 5.12 muestra entonces que si completamos $\sigma(\mathcal{A})$ con respecto a μ, obtenemos la σ-álgebra generada por $\mathcal{A}$ y los conjuntos de medida exterior cero, es decir, recuperamos la medida μ definida sobre $\Im$.

5.4. Teorema de clases monótonas

Es muy frecuente encontrar un problema del siguiente tipo: Se tiene un conjunto $\mathbb{F}$, un álgebra $\mathcal{A}$ de subconjuntos de $\mathbb{F}$ y la σ-álgebra $\sigma(\mathcal{A})$ generada por $\mathcal{A}$ y se quiere demostrar que una cierta propiedad es válida para todo elemento de $\sigma(\mathcal{A})$. Un método para resolver este problema consiste en demostrar que la familia $\mathcal{H}$ de subconjuntos de $\mathbb{F}$ que tienen la propiedad deseada es una σ-álgebra que contiene a todos los elementos de $\mathcal{A}$, de manera que entonces contiene a la σ-álgebra generada por $\mathcal{A}$. Sin embargo, en ocasiones puede resultar sumamente complicado demostrar que efectivamente $\mathcal{H}$ forma una σ-álgebra, sobre todo la demostración de que $\mathcal{H}$ es cerrada bajo uniones o intersecciones finitas. Para salvar esta dificultad, se tienen, afortunadamente, los resultados que se exponen en esta sección, los cuales son conocidos como teoremas de clases monótonas, cuyo origen se deben al trabajo de Dynkin.

Definición 5.12. *Sea $\mathbb{F}$ un conjunto y $\mathcal{G}$ una familia de subconjuntos de $\mathbb{F}$. Diremos que:*

(i) *$\mathcal{G}$ es cerrada bajo complementos si $A^c \in \mathcal{G}$ para cualquier $A \in \mathcal{G}$.*
(ii) *$\mathcal{G}$ es cerrada bajo diferencias propias si $B - A \in \mathcal{G}$ para cualquier pareja $A, B \in \mathcal{G}$ tal que $A \subset B$.*
(iii) *$\mathcal{G}$ es cerrada bajo uniones (resp. intersecciones) finitas si $\bigcup_{j=1}^{n} A_j \in \mathcal{G}$ (resp. $\bigcap_{j=1}^{n} A_j \in \mathcal{G}$) para cualquier colección finita $A_1, A_2, \ldots, A_n$ de elementos de $\mathcal{G}$.*
(iv) *$\mathcal{G}$ es cerrada bajo uniones (resp. intersecciones) monótonas si $\bigcup_{j=1}^{\infty} A_j \in \mathcal{G}$ (resp. $\bigcap_{j=1}^{\infty} A_j \in \mathcal{G}$) para cualquier sucesión creciente (resp. decreciente) $(A_n)_{n\in\mathbb{N}}$, de elementos de $\mathcal{G}$.*

Definición 5.13. *Sea $\mathbb{F}$ un conjunto y $\mathcal{M}$ una familia de subconjuntos de $\mathbb{F}$. Se dice que $\mathcal{M}$ es una clase monótona si es cerrada bajo uniones e intersecciones monótonas.*

Definición 5.14. *Dado un conjunto $\mathbb{F}$ y una familia arbitraria de clases monótonas de subconjuntos de $\mathbb{F}$, se define la intersección de esas clases monótonas como la familia de conjuntos que pertenecen a todas ellas.*

Se puede ver fácilmente que la intersección de clases monótonas, de subconjuntos de un conjunto $\mathbb{F}$, forma una clase monótona.

Definición 5.15. *Dada una colección $\mathcal{A}$ de subconjuntos de un conjunto $\mathbb{F}$, se define la clase monótona generada por $\mathcal{A}$ como la intersección de todas las clases monótonas que contienen a todos los elementos de $\mathcal{A}$ y se le denota por $\mathcal{M}(\mathcal{A})$.*

Obsérvese que la definición anterior es consistente pues dada cualquier colección $\mathcal{A}$ de subconjuntos de un conjunto $\mathbb{F}$ existe por lo menos una clase monótona que contiene a todos los conjuntos de $\mathcal{A}$, a saber, la clase monótona formada por todos los subconjuntos de $\mathbb{F}$.

Obsérvese también que la clase monótona generada por una familia $\mathcal{A}$ de subconjuntos de un conjunto $\mathbb{F}$ es la más pequeña clase monótona de subconjuntos de $\mathbb{F}$ que contiene a todos los elementos de $\mathcal{A}$.

Teorema 5.7. *Sea $\mathbb{F}$ un conjunto y $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$, entonces la clase monótona generada por $\mathcal{A}$ sigue siendo un álgebra.*

Demostración

Para demostrar que $\mathcal{M}(\mathcal{A})$ es cerrada bajo complementos, sea:

$\mathcal{C} = \{A \in \mathcal{M}(\mathcal{A}) : A^c \in \mathcal{M}(\mathcal{A})\}$.

$\mathcal{C}$ es entonces una clase monótona. En efecto, si $A_1 \subset A_2 \cdots$ es una sucesión de elementos de $\mathcal{C}$ entonces, para cualquier $n \in \mathbb{N}$, se tiene $A_n \in \mathcal{M}(\mathcal{A})$ y $A_n^c \in \mathcal{M}(\mathcal{A})$, así que $\bigcup_n A_n \in \mathcal{M}(\mathcal{A})$ y $(\bigcup_n A_n)^c = \bigcap_n A_n^c \in \mathcal{M}(\mathcal{A})$, por lo tanto, $\bigcup_n A_n \in \mathcal{C}$. De la misma manera se demuestra que $\mathcal{C}$ es cerrada bajo intersecciones monótonas.

Obviamente $\mathcal{C}$ contiene a $\mathcal{A}$, de manera que entonces se puede concluir que $\mathcal{C}$ contiene a $\mathcal{M}(\mathcal{A})$, lo cual prueba que $\mathcal{M}(\mathcal{A})$ es cerrada bajo complementos.

Para demostrar que $\mathcal{M}(\mathcal{A})$ es cerrada bajo intersecciones finitas, sea:

$\mathcal{C}_1 = \{A \in \mathcal{M}(\mathcal{A}) : A \cap B \in \mathcal{M}(\mathcal{A})$ para cualquier $B \in \mathcal{A}\}$.

$\mathcal{C}_1$ es entonces una clase monótona. En efecto, si $A_1 \subset A_2 \cdots$ es una sucesión de elementos de $\mathcal{C}_1$ entonces, para cualquier $n \in \mathbb{N}$ y $B \in \mathcal{A}$, se tiene $A_n \in \mathcal{M}(\mathcal{A})$ y $A_n \cap B \in \mathcal{M}(\mathcal{A})$, así que $\bigcup_n A_n \in \mathcal{M}(\mathcal{A})$ y $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B) \in \mathcal{M}(\mathcal{A})$, por lo tanto, $\bigcup_n A_n \in \mathcal{C}_1$. De la misma manera se demuestra que $\mathcal{C}_1$ es cerrada bajo intersecciones monótonas.

Obviamente $\mathcal{C}_1$ contiene a $\mathcal{A}$, de manera que entonces se puede concluir que $\mathcal{C}_1$ contiene a $\mathcal{M}(\mathcal{A})$, es decir, para cualquier $A \in \mathcal{M}(\mathcal{A})$ y $B \in \mathcal{A}$, se tiene $A \cap B \in \mathcal{M}(\mathcal{A})$.

Sea ahora:

$\mathcal{C}_2 = \{A \in \mathcal{M}(\mathcal{A}) : A \cap B \in \mathcal{M}(\mathcal{A})$ para cualquier $B \in \mathcal{M}(\mathcal{A})\}$.

$\mathcal{C}_2$ es entonces una clase monótona. En efecto, si $A_1 \subset A_2 \cdots$ es una sucesión de elementos de $\mathcal{C}_2$ entonces, para cualquier $n \in \mathbb{N}$ y $B \in \mathcal{M}(\mathcal{A})$, se tiene $A_n \in \mathcal{M}(\mathcal{A})$ y $A_n \cap B \in \mathcal{M}(\mathcal{A})$, así que $\bigcup_n A_n \in \mathcal{M}(\mathcal{A})$ y $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B) \in \mathcal{M}(\mathcal{A})$, por lo tanto, $\bigcup_n A_n \in \mathcal{C}_2$. De la misma manera se demuestra que $\mathcal{C}_2$ es cerrada bajo intersecciones monótonas.

Como $\mathcal{C}_1$ contiene a $\mathcal{M}(\mathcal{A})$ se tiene que $\mathcal{C}_2$ contiene a $\mathcal{A}$, de manera que entonces se puede concluir que $\mathcal{C}_2$ contiene a $\mathcal{M}(\mathcal{A})$, es decir, para cualesquiera $A, B \in \mathcal{M}(\mathcal{A})$, se tiene $A \cap B \in \mathcal{M}(\mathcal{A})$.

■

Corolario 5.3 (Teorema de clases monótonas para álgebras). *Sea $\mathbb{F}$ un conjunto y $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$, entonces $\sigma(A) = \mathcal{M}(\mathcal{A})$.*

El teorema 5.7 puede ser insuficiente pues en ocasiones únicamente se puede demostrar inmediatamente que la propiedad que se quiere probar como válida para cualquier elemento de una σ-álgebra se cumple para una familia de conjuntos que la generan y que es cerrada sólo bajo intersecciones finitas (por ejemplo, la familia formada por todos los intervalos de números reales).

Definición 5.16. *Sea $\mathbb{F}$ un conjunto y $\mathcal{P}$ una familia de subconjuntos de $\mathbb{F}$. Se dice que $\mathcal{P}$ es un π-sistema si es cerrada bajo intersecciones finitas.*

Definición 5.17. *Sea $\mathbb{F}$ un conjunto y $\mathcal{D}$ una familia de subconjuntos de $\mathbb{F}$. Se dice que $\mathcal{D}$ es un d-sistema si $\mathbb{F} \in \mathcal{D}$ y es cerrada bajo diferencias propias y uniones monótonas.*

Definición 5.18. *Dado un conjunto $\mathbb{F}$ y una familia arbitraria de d-sistemas de subconjuntos de $\mathbb{F}$, se define la intersección de esos d-sistemas como la familia de conjuntos que pertenecen a todas ellas.*

Se puede ver fácilmente que la intersección de d-sistemas, de subconjuntos de un conjunto $\mathbb{F}$, forma un d-sistema.

Definición 5.19. *Dada una colección $\mathcal{G}$ de subconjuntos de un conjunto $\mathbb{F}$, se define el d-sistema generado por $\mathcal{G}$ como la intersección de todos los d-sistemas que contienen a todos los elementos de $\mathcal{G}$ y se le denota por $d(\mathcal{G})$.*

Obsérvese que la definición anterior es consistente pues dada cualquier colección $\mathcal{P}$ de subconjuntos de un conjunto $\mathbb{F}$ existe por lo menos un d-sistema que contiene a todos los conjuntos de $\mathcal{P}$, a saber, el d-sistema formada por todos los subconjuntos de $\mathbb{F}$.

Obsérvese también que el d-sistema generado por una familia $\mathcal{P}$ de subconjuntos de un conjunto $\mathbb{F}$ es el más pequeño d-sistema de subconjuntos de $\mathbb{F}$ que contiene a todos los elementos de $\mathcal{P}$.

Teorema 5.8. *Sea $\mathbb{F}$ un conjunto y $\mathcal{P}$ un π-sistema de subconjuntos de $\mathbb{F}$, entonces el d-sistema generado por $\mathcal{P}$ es un π-sistema.*

Demostración

Sea:

$\mathcal{C}_1 = \{A \in d(\mathcal{P}) : A \cap B \in d(\mathcal{P}) \text{ para cualquier } B \in \mathcal{P}\}$.

$\mathcal{C}_1$ es entonces un d-sistema. En efecto, si $A_1 \subset A_2 \cdots$ es una sucesión de elementos de $\mathcal{C}_1$ entonces, para cualquier $n \in \mathbb{N}$ y $B \in \mathcal{P}$, se tiene $A_n \in d(\mathcal{P})$ y $A_n \cap B \in d(\mathcal{P})$, así que $\bigcup_n A_n \in d(\mathcal{P})$ y $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B) \in d(\mathcal{P})$, por lo tanto, $\bigcup_n A_n \in \mathcal{C}_1$. Sean ahora $A, C \in \mathcal{C}_1$ tales que $A \subset C$, entonces $A, C, A \cap B, C \cap B \in d(\mathcal{P})$ y $A \cap B \subset C \cap B$ para cualquier $B \in \mathcal{P}$, así que $C - A \in d(\mathcal{P})$ y $(C - A) \cap B = C \cap B - A \cap B \in d(\mathcal{P})$, por lo tanto, $C - A \in \mathcal{C}_1$. Finalmente, es obvio que $F \in \mathcal{C}_1$.

Obviamente $\mathcal{C}_1$ contiene a $\mathcal{P}$, de manera que entonces se puede concluir que $\mathcal{C}_1$ contiene a $d(\mathcal{P})$, es decir, para cualquier $A \in d(\mathcal{P})$ y $B \in \mathcal{P}$, se tiene $A \cap B \in d(\mathcal{P})$.

Sea ahora:

$\mathcal{C}_2 = \{A \in d(\mathcal{P}) : A \cap B \in d(\mathcal{P}) \text{ para cualquier } B \in d(\mathcal{P})\}$.

$\mathcal{C}_2$ es entonces un d-sistema. En efecto, si $A_1 \subset A_2 \cdots$ es una sucesión de elementos de $\mathcal{C}_2$ entonces, para cualquier $n \in \mathbb{N}$ y $B \in d(\mathcal{P})$, se tiene $A_n \in d(\mathcal{P})$ y $A_n \cap B \in d(\mathcal{P})$, así que $\bigcup_n A_n \in d(\mathcal{P})$ y $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B) \in d(\mathcal{P})$, por lo tanto, $\bigcup_n A_n \in \mathcal{C}_2$. Sean ahora $A, C \in \mathcal{C}_2$ tales que $A \subset C$, entonces $A, C, A \cap B, C \cap B \in d(\mathcal{P})$ y $A \cap B \subset C \cap B$ para cualquier $B \in d(\mathcal{P})$, así que $C - A \in d(\mathcal{P})$ y $(C - A) \cap B = C \cap B - A \cap B \in d(\mathcal{P})$, por lo tanto, $C - A \in \mathcal{C}_2$. Finalmente, es obvio que $\Omega \in \mathcal{C}_2$.

Como $\mathcal{C}_1$ contiene a $d(\mathcal{P})$, se tiene que $\mathcal{C}_2$ contiene a $\mathcal{P}$, de manera que entonces se puede concluir que $\mathcal{C}_2$ contiene a $d(\mathcal{P})$, es decir, para cualesquiera $A, B \in d(\mathcal{P})$, se tiene $A \cap B \in d(\mathcal{P})$.

∎

Corolario 5.4 (Teorema de clases monótonas para pi sistemas). *Sea $\mathbb{F}$ un conjunto y $\mathcal{P}$ un π-sistema de subconjuntos de $\mathbb{F}$, entonces $d(\mathcal{P}) = \sigma(\mathcal{P})$.*

5.5. Unicidad de la extensión de una medida

Teorema 5.9. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y μ_1 y μ_2 dos medidas definidas sobre $\sigma(\mathcal{A})$, tales que $\mu_1(A) = \mu_2(A)$ para cualquier $A \in \mathcal{A}$. Supongamos además que existe una sucesión creciente $(F_n)_{n\in\mathbb{N}}$ de elementos de $\sigma(\mathcal{A})$ tal que $\bigcup_{n=1}^{\infty} F_n = \mathbb{F}$ y, para cualquier $n \in \mathbb{N}$, $\mu_1(F_n) = \mu_2(F_n) < \infty$ y $\mu_1(A \cap F_n) = \mu_2(A \cap F_n)$ para cualquier $A \in \mathcal{A}$. Entonces $\mu_1(A) = \mu_2(A)$ para cualquier $A \in \sigma(\mathcal{A})$.*

Demostración

Para cada $n \in \mathbb{N}$, definamos:

$\mathcal{H}_n = \{A \in \sigma(\mathcal{A}) : \mu_1(A \cap F_n) = \mu_2(A \cap F_n)\}$.

Por hipótesis, $A \in \mathcal{H}_n$ para cualquier $A \in \mathcal{A}$.

Si $(A_m)_{m\in\mathbb{N}}$ es una sucesión creciente de elementos de $\mathcal{H}_n$, entonces:

$\mu_1\left(\left(\bigcup_{m=1}^{\infty} A_m\right)\cap F_n\right) = \mu_1\left(\bigcup_{m=1}^{\infty}\left(A_m\cap F_n\right)\right) = \text{lím}_{n\rightsquigarrow\infty}\,\mu_1\left(A_m\cap F_n\right)$

$= \text{lím}_{n\rightsquigarrow\infty}\,\mu_2\left(A_m\cap F_n\right) = \mu_2\left(\bigcup_{m=1}^{\infty}\left(A_m\cap F_n\right)\right) = \mu_2\left(\left(\bigcup_{m=1}^{\infty} A_m\right)\cap F_n\right)$.

Así que $\bigcup_{m=1}^{\infty} A_m \in \mathcal{H}_n$.

Si $(A_m)_{m\in\mathbb{N}}$ es una sucesión decreciente de elementos de $\mathcal{H}_n$, entonces:

$\mu_1\left(\left(\bigcap_{m=1}^{\infty} A_m\right)\cap F_n\right) = \mu_1\left(\bigcap_{m=1}^{\infty}\left(A_m\cap F_n\right)\right) = \text{lím}_{n\rightsquigarrow\infty}\,\mu_1\left(A_m\cap F_n\right)$

$= \text{lím}_{n\rightsquigarrow\infty}\,\mu_2\left(A_m\cap F_n\right) = \mu_2\left(\bigcap_{m=1}^{\infty}\left(A_m\cap F_n\right)\right) = \mu_2\left(\left(\bigcap_{m=1}^{\infty} A_m\right)\cap F_n\right)$.

Así que$\bigcap_{m=1}^{\infty} A_m \in \mathcal{H}_n$.

$\mathcal{H}_n$ es entonces una clase monótona que contiene a los elementos de $\mathcal{A}$, así que $\mathcal{H}_n$ contiene a $\sigma(\mathcal{A}) = \mathcal{M}(\mathcal{A})$.

Finalmente, si $A\in\sigma(\mathcal{A})$, se tiene:

$\mu_1(A) = \text{lím}_{n\rightsquigarrow\infty}\,\mu_1\left(A\cap F_n\right) = \text{lím}_{n\rightsquigarrow\infty}\,\mu_2\left(A\cap F_n\right) = \mu_2(A)$.

■

Corolario 5.5. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y μ_1 y μ_2 dos medidas finitas definidas sobre $\sigma(\mathcal{A})$, tales que $\mu_1(A) = \mu_2(A)$ para cualquier $A\in\mathcal{A}$. Entonces $\mu_1(A) = \mu_2(A)$ para cualquier $A\in\sigma(\mathcal{A})$.*

Corolario 5.6. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y μ_1 y μ_2 dos medidas definidas sobre $\sigma(\mathcal{A})$, tales que $\mu_1(A) = \mu_2(A)$ para cualquier $A\in\mathcal{A}$. Supongamos además que existe una sucesión $(F_n)_{n\in\mathbb{N}}$ de elementos de $\sigma(\mathcal{A})$, ajenos por parejas, tal que $\bigcup_{n=1}^{\infty} F_n = \mathbb{F}$ y, para cualquier $n\in\mathbb{N}$, $\mu_1(F_n) = \mu_2(F_n) < \infty$ y $\mu_1(A\cap F_n) = \mu_2(A\cap F_n)$ para cualquier $A\in\mathcal{A}$. Entonces $\mu_1(A) = \mu_2(A)$ para cualquier $A\in\sigma(\mathcal{A})$.*

Corolario 5.7. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y μ_1 y μ_2 dos medidas definidas sobre $\sigma(\mathcal{A})$, tales que $\mu_1(A) = \mu_2(A)$ para cualquier $A\in\mathcal{A}$. Supongamos además que existe una sucesión $(F_n)_{n\in\mathbb{N}}$ de elementos de $\mathcal{A}$ tal que $\bigcup_{n=1}^{\infty} F_n = \mathbb{F}$ y $\mu_1(F_n) = \mu_2(F_n) < \infty$ para cualquier $n\in\mathbb{N}$. Entonces $\mu_1(A) = \mu_2(A)$ para cualquier $A\in\sigma(\mathcal{A})$.*

Demostración

Para cada $n\in\mathbb{N}$, definamos $E_n = \bigcup_{k=1}^{n} F_k$, entonces $E_n\in\mathcal{A}$, así que $\mu_1(A\cap F_n) = \mu_2(A\cap F_n)$ para cualquier $A\in\mathcal{A}$. Además $\mu_1(E_n) = \mu_2(E_n) \leq \sum_{k=1}^{n}\mu_2(F_n) < \infty$ para cualquier $n\in\mathbb{N}$ y $\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} F_n = \mathbb{F}$. Así que, por la proposición anterior, $\mu_1(A) = \mu_2(A)$ para cualquier $A\in\sigma(\mathcal{A})$.

■

Teorema 5.10. *Sea $\mathbb{F}$ un conjunto, $\mathcal{P}$ un π-sistema de subconjuntos de $\mathbb{F}$ y μ_1 y μ_2 dos medidas definidas sobre $\sigma(\mathcal{P})$, tales que $\mu_1(A) = \mu_2(A)$ para cualquier $A\in\mathcal{P}$. Supongamos* *además que existe una sucesión creciente $(F_n)_{n\in\mathbb{N}}$ de elementos de $\sigma(\mathcal{P})$ tales que $\bigcup_{n=1}^{\infty} F_n = \mathbb{F}$ y, para cualquier $n \in \mathbb{N}$, $\mu_1(F_n) = \mu_2(F_n) < \infty$ y $\mu_1(A \cap F_n) = \mu_2(A \cap F_n)$ para cualquier $A \in \mathcal{P}$. Entonces $\mu_1(A) = \mu_2(A)$ para cualquier $A \in \sigma(\mathcal{P})$.*

Demostración

Para cada $n \in \mathbb{N}$, definamos:

$\mathcal{H}_n = \{A \in \sigma(\mathcal{P}) : \mu_1(A \cap F_n) = \mu_2(A \cap F_n)\}$.

Obviamente, $\mathbb{F} \in \mathcal{H}_n$ y, por hipótesis, $A \in \mathcal{H}_n$ para cualquier $A \in \mathcal{P}$.

Si $A, B \in \mathcal{H}_n$ y $A \subset B$, entonces:

$\mu_1((B - A) \cap F_n) = \mu_1(B \cap F_n - A \cap F_n) = \mu_1(B \cap F_n) - \mu_1(A \cap F_n)$

$= \mu_2(B \cap F_n) - \mu_2(A \cap F_n) = \mu_2(B \cap F_n - A \cap F_n) = \mu_2((B - A) \cap F_n)$.

Así que $B - A \in \mathcal{H}_n$.

Si $(A_m)_{m\in\mathbb{N}}$ es una sucesión creciente de elementos de $\mathcal{H}_n$, entonces:

$\mu_1\left(\left(\bigcup_{m=1}^{\infty} A_m\right) \cap F_n\right) = \mu_1\left(\bigcup_{m=1}^{\infty} (A_m \cap F_n)\right) = \lim_{n\rightsquigarrow\infty} \mu_1\left(A_m \cap F_n\right)$.

$= \lim_{n\rightsquigarrow\infty} \mu_2\left(A_m \cap F_n\right) = \mu_2\left(\bigcup_{m=1}^{\infty} (A_m \cap F_n)\right) = \mu_2\left(\left(\bigcup_{m=1}^{\infty} A_m\right) \cap F_n\right)$

Así que $\bigcup_{m=1}^{\infty} A_m \in \mathcal{H}_n$.

$\mathcal{H}_n$ es entonces un d-sistema que contiene a los elementos de $\mathcal{P}$, así que $\mathcal{H}_n$ contiene a $\sigma(\mathcal{P}) = d(\mathcal{P})$.

Finalmente, si $A \in \sigma(\mathcal{P})$, se tiene:

$\mu_1\left(A\right) = \lim_{n\rightsquigarrow\infty} \mu_1\left(A \cap F_n\right) = \lim_{n\rightsquigarrow\infty} \mu_2\left(A \cap F_n\right) = \mu_2\left(A\right)$.

■

COROLARIO 5.8. *Sea $\mathbb{F}$ un conjunto, $\mathcal{P}$ un π-sistema de subconjuntos de $\mathbb{F}$ y μ_1 y μ_2 dos medidas finitas definidas sobre $\sigma(\mathcal{P})$, tales que $\mu_1(\mathbb{F}) = \mu_2(\mathbb{F})$ y $\mu_1(A) = \mu_2(A)$ para cualquier $A \in \mathcal{P}$. Entonces $\mu_1(A) = \mu_2(A)$ para cualquier $A \in \sigma(\mathcal{P})$.*

COROLARIO 5.9. *Sea $\mathbb{F}$ un conjunto, $\mathcal{P}$ un π-sistema de subconjuntos de $\mathbb{F}$ y μ_1 y μ_2 dos medidas definidas sobre $\sigma(\mathcal{P})$, tales que $\mu_1(A) = \mu_2(A)$ para cualquier $A \in \mathcal{P}$. Supongamos además que existe una sucesión $(F_n)_{n\in\mathbb{N}}$ de elementos de $\sigma(\mathcal{P})$, ajenos por parejas, tal que $\bigcup_{n=1}^{\infty} F_n = \mathbb{F}$ y, para cualquier $n \in \mathbb{N}$, $\mu_1(F_n) = \mu_2(F_n) < \infty$ y $\mu_1(A \cap F_n) = \mu_2(A \cap F_n)$ para cualquier $A \in \mathcal{P}$. Entonces $\mu_1(A) = \mu_2(A)$ para cualquier $A \in \sigma(\mathcal{P})$.*

COROLARIO 5.10. *Sea $\mathbb{F}$ un conjunto, $\mathcal{P}$ un π-sistema de subconjuntos de $\mathbb{F}$ y μ_1 y μ_2 dos medidas definidas sobre $\sigma(\mathcal{P})$, tales que $\mu_1(A) = \mu_2(A)$ para cualquier $A \in \mathcal{P}$. Supongamos además que existe una sucesión $(F_n)_{n\in\mathbb{N}}$ de elementos de $\mathcal{P}$ tales que $\bigcup_{n=1}^{\infty} F_n = \mathbb{F}$ y $\mu_1(F_n) = \mu_2(F_n) < \infty$, para cualquier $n \in \mathbb{N}$. Entonces $\mu_1(A) = \mu_2(A)$ para cualquier $A \in \sigma(\mathcal{P})$.*

Demostración

Para cada $n \in \mathbb{N}$, definamos $E_n = \bigcup_{k=1}^{n} F_k$.

Por la proposición 2.11, se tiene $\mu_1(E_n) = \mu_2(E_n)$.

Además, $\mu_1(E_n) = \mu_2(E_n) \leq \sum_{k=1}^{n} \mu_2(F_n) < \infty$ para cualquier $n \in \mathbb{N}$ y:

$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} F_n = \mathbb{F}$.

Si $A \in \mathcal{P}$, nuevamente por proposición 2.11, se tiene:

$\mu_1(A \cap E_n) = \mu_1(A \cap (\bigcup_{k=1}^{n} F_k)) = \mu_1(\bigcup_{k=1}^{n} (A \cap F_k))$

$= \mu_2(\bigcup_{k=1}^{n} (A \cap F_k)) = \mu_2(A \cap (\bigcup_{k=1}^{n} F_k)) = \mu_2(A \cap E_n)$.

Así que, por la proposición anterior, $\mu_1(A) = \mu_2(A)$ para cualquier $A \in \sigma(\mathcal{P})$.

■

5.6. Medidas con signo

DEFINICIÓN 5.20. *Se dice que una función $\mu : \Im \mapsto \mathbb{R} \cup \{-\infty, \infty\}$ es σ-aditiva si μ toma a lo más uno de los valores $+\infty$ y $-\infty$, y, dada cualquier familia numerable, $A_1, A_2, \ldots$, de elementos de $\Im$ tal que $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces $\mu\left(\bigcup\limits_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$, donde la serie $\sum_{k=1}^{\infty} \mu(A_k)$ converge absolutamente cuando $\mu\left(\bigcup\limits_{k=1}^{\infty} A_k\right) \in \mathbb{R}$.*

DEFINICIÓN 5.21. *e dice que $\nu : \Im \mapsto \mathbb{R} \cup \{-\infty, \infty\}$ es una medida con signo si se cumplen las siguientes condiciones*

(i) *$\nu(\emptyset) = 0$.*
(ii) *ν es σ-aditiva.*

DEFINICIÓN 5.22. *Sea ν una medida con signo y $A \in \Im$. Se dice que A es un conjunto positivo (resp. negativo) con respecto a ν si $\nu(E) \geq 0$ (resp. $\nu(E) \leq 0$) para cualquier conjunto medible $E \subset A$.*

PROPOSICIÓN 5.13. *La unión de una colección finita o infinita numerable de conjuntos positivos es un conjunto positivo.*

Demostración

Sea $\{A_n\}$ una colección finita o infinita numerable de conjuntos positivos, $A = \bigcup_n A_n$, E un conjunto medible contenido en A y $E'_n = E \cap A_n \cap A_1^c \cap \cdots \cap A_{n-1}^c$. Entonces los conjuntos E'_n son ajenos por parejas, $E = \bigcup_n E'_n$ y, como $E'_n \subset A_n$, $\nu(E'_n) \geq 0$; así que $\nu(E) \geq 0$.

■

TEOREMA 5.11. *Sea ν una medida con signo y E un conjunto medible tal que $0 < \nu(E) < \infty$. Entonces existe un conjunto positivo $A \subset E$ tal que $0 < \nu(A) < \infty$ y $C = E - A$ es un conjunto negativo.*

Demostración

La idea consiste en irle quitando al conjunto E conjuntos medibles de medida negativa hasta legar a un conjunto con la propiedad deseada.

Obsérvese que si B es un subconjunto de E de medida menor o igual a 0, entonces, como $E = B \cup (E - B)$, se tiene $0 < \nu(E - B) < \infty$.

Por otra parte, E no contiene subconjuntos de medida $-\infty$ pues si B fuera un subconjunto medible de E de medida $-\infty$, se tendría $E = B \cup (E - B)$, con $\nu(E - B) < \infty$, así que $\nu(E) = -\infty$.

E tampoco contiene subconjuntos de medida ∞ pues si B fuera un subconjunto medible de E de medida ∞, se tendría $E = B \cup (E - B)$, con $\nu(E - B) > -\infty$, así que $\nu(E) = \infty$.

Ahora obsérvese que dada $\varepsilon > 0$, existe un conjunto medible $B \subset E$ tal que $\nu(B) \leq 0$ y $E - B$ no contiene ningún subconjunto de medida menor que $-\varepsilon$. En efecto, si no existe un conjunto medible $B_1 \subset E$ tal que $\nu(B_1) < -\varepsilon$, tomemos $B = \emptyset$; en otro caso, si no existe un conjunto medible $B_2 \subset E - B_1$ tal que $\nu(B_2) < -\varepsilon$, tomemos $B = B_1$; en otro caso, si no existe un conjunto medible $B_3 \subset E - B_1 \cup B_2$ tal que $\nu(B_3) < -\varepsilon$, tomemos $B = B_1 \cup B_2$. Continuando con este procedimiento, se llega, en un número finito de pasos, a la obtención de un conjunto B con la propiedad deseada, pues si el procedimiento continuara indefinidamente obtendríamos una colección infinita de conjuntos medibles contenidos en E, ajenos por parejas, cuya unión sería un subconjunto medible de E de medida $-\infty$.

Tomemos entonces un conjunto medible $B_1 \subset E$ tal que $\nu(B_1) \leq 0$ y $E - B_1$ no contiene ningún subconjunto de medida menor que -1. Inductivamente podemos obtener una colección infinita $B_1, B_2, \ldots$ de conjuntos medibles contenidos en E, ajenos por parejas y tales que $E - \cup_{k=1}^{n} B_k$ no contiene ningún subconjunto de medida menor que $-\frac{1}{n}$. Definiendo $B = \cup_{k=1}^{\infty} B_k$, se tiene que $\nu(B) \leq 0$ y $A = E - B$ no contiene ningún subconjunto medible de medida negativa, es decir, es un conjunto positivo tal que $0 < \nu(A) < \infty$.

Hemos demostrado hasta aquí que existe un conjunto positivo $A \subset E$ tal que $0 < \nu(A) < \infty$.

Sea A_1 un conjunto positivo cualquiera tal que $A_1 \subset E$ y $0 < \nu(A_1) < \infty$.

Demostremos ahora que dada $\varepsilon > 0$, existe un conjunto positivo $A \subset E$ tal que $0 < \nu(A) < \infty$ y $E - A$ no contiene ningún conjunto positivo de medida mayor que ε. En efecto, si no existe un conjunto positivo $A_2 \subset E - A_1$ tal que $\nu(A_2) > 1$, tomemos $A = A_1$; en otro caso, si no existe un conjunto positivo $A_3 \subset E - A_1 \cup A_2$ tal que $\nu(A_3) > 1$, tomemos $A = A_1 \cup A_2$; en otro caso, si no existe un conjunto positivo $A_4 \subset E - A_1 \cup A_2 \cup A_3$ tal que $\nu(A_3) > 1$, tomemos $A = A_1 \cup A_2 \cup A_3$. Continuando con este procedimiento, se llega, en un número finito de pasos, a la obtención de un conjunto A con la propiedad deseada, pues si el procedimiento continuara indefinidamente obtendríamos una colección infinita de conjuntos medibles contenidos en E, ajenos por parejas, cuya unión sería un subconjunto medible de E de medida ∞.

Tomemos entonces un conjunto positivo $A_1 \subset E$ tal que $0 < \nu(A_1) < \infty$ y $E - A_1$ no contiene ningún conjunto positivo de medida mayor que 1. Inductivamente podemos obtener una colección infinita $A_1, A_2, \ldots$ de conjuntos medibles contenidos en E, ajenos por parejas y tales que $E - \cup_{k=1}^{n} B_k$ no contiene ningún subconjunto de medida menor que $-\frac{1}{n}$. Definiendo $B = \cup_{k=1}^{\infty} B_k$, se tiene que $\nu(B) \leq 0$ y $A = E - B$ no contiene ningún subconjunto medible de medida negativa, es decir, es un conjunto positivo tal que $0 < \nu(A) < \infty$.

Si E es un conjunto positivo, definimos $A = E$ y termina la demostración. En otro caso, existe un conjunto medible $B \subset E$ tal que $\nu(B) < 0$. Sea n_1 el más pequeño entero positivo para el cual existe un conjunto medible $B \subset E$ tal que $\nu(B) < -\frac{1}{n_1}$, B_1 un conjunto medible con esa propiedad y $E_1 = E - B_1$. Entonces se tiene $0 < \nu(E_1) < \infty$.

Si E_1 es un conjunto positivo, definimos $A = E_1$ y termina la demostración. En otro caso, existe un conjunto medible $B \subset E_1$ tal que $\nu(B) < 0$. Sea n_2 el más pequeño entero positivo para el cual existe un conjunto medible $B \subset E_1$ tal que $\nu(B) < -\frac{1}{n_2}$, B_2 un conjunto medible con esa propiedad y $E_2 = E_1 - B_2$. Entonces se tiene $0 < \nu(E_2) < \infty$.

Continuando con este proceso, o bien se llega al resultado deseado en un número finito de pasos, o bien se obtiene una sucesión de enteros positivos (n_k) y sucesión (B_k) de subconjuntos medibles de E tales que, para cualquier $k \in \mathbf{N}$, $\nu(B_k) < -\frac{1}{n_k}$ y $\nu(B) \geq -\frac{1}{n_k - 1}$ para cualquier subconjunto medible $B \subset E - \bigcup_{j=1}^{k} B_j$. En este último caso, sea $A = E - \bigcup_{j=1}^{\infty} B_j$. Entonces se tiene $0 < \nu(A) < \infty$.

Obsérvese ahora que necesariamente se tiene $n_k \leq n_{k+1}$. Además, cada término n_k se repite, a lo más, un número finito de veces pues de otra manera la unión de los correspondientes conjuntos B_k sería un conjunto medible de E de medida $-\infty$. Por lo tanto, $\lim_{k \rightsquigarrow \infty} n_k = \infty$.

Finalmente, como $A \subset E - \bigcup_{j=1}^{k} B_j$ para cualquier $k \in \mathbf{N}$, entonces, dado cualquier subconjunto medible $B \subset A$ se tiene $\nu(B) \geq -\frac{1}{n_k - 1}$ para cualquier $k \in \mathbb{N}$. Por lo tanto, tomando límite cuando $k \rightsquigarrow \infty$, se tiene $\nu(B) \geq 0$, así que A es un conjunto positivo.

■

Teorema 5.12 (**Teorema de descomposición de Hahn**). *Sea ν una medida con signo. Entonces existe un conjunto positivo A y un conjunto negativo B tales que $A \cap B = \emptyset$ y $\mathbb{F} = A \cup B$.*

Demostración

La idea consiste en lo siguiente:

Si ν no toma el valor $+\infty$, se trata de encontrar un conjunto positivo de medida máxima. El complemento de ese conjunto es entonces, necesariamente, un conjunto negativo.

Si ν no toma el valor $-\infty$, se trata de encontrar un conjunto negativo de medida mínima. El complemento de ese conjunto es entonces, necesariamente, un conjunto positivo.

Supongamos que ν no toma el valor $+\infty$.

Sea $\lambda = \sup\{\nu(A) : A$ es un conjunto positivo con respecto a $\nu\}$.

Como el conjunto vacío es positivo, se tiene $\lambda \geq 0$.

Sea $\{A_n\}$ una sucesión de conjuntos positivos tales que $\lambda = \lim_{n \rightsquigarrow \infty} \nu(A_n)$ y sean $A = \bigcup_n A_n$ y $B = A^c$.

A es un conjunto positivo por ser la unión numerable de conjuntos positivos.

Por ser A un conjunto positivo, se tiene $\lambda \geq \nu(A)$. Por otra parte, $A - A_n \subset A$ para cualquier n, así que $\nu(A - A_n) \geq 0$. Así que, $\nu(A) = v(A_n) + \nu(A - A_n) \geq v(A_n)$ para cualquier n. Por lo tanto, tomando límite cuando $n \to \infty$, se tiene $\nu(A) \geq \lambda$. Se concluye entonces que $\lambda = \nu(A)$, así que $\lambda < \infty$.

Para probar que B es un conjunto negativo, supongamos que existe un conjunto medible $E \subset B$ tal que $\nu(E) > 0$, entonces, por la proposición 5.11, existe un conjunto positivo $D \subset E$ tal que $\nu(D) > 0$. Por lo tanto, $A \cup D$ es un conjunto positivo y como A y D son ajenos, se tiene $\lambda \geq \nu(A \cup D) = \nu(A) + \nu(D) = \lambda + \nu(D) > \lambda$, lo cual es una contradicción.

■

Definición 5.23. *Sea ν una medida con signo. Entonces una pareja de conjuntos medibles (A, B) tales que A es positivo, B es negativo, $A \cap B = \emptyset$ y $\mathbb{F} = A \cup B$, es llamada una descomposición de Hahn para ν.*

Teorema 5.13. *Sea $\nu : (\mathbb{F}, \Im) \mapsto \mathbb{R} \cup \{-\infty, \infty\}$ una medida con signo, entonces existen dos medidas ν^+ y ν^- sobre $(\mathbb{F}, \Im)$ con las siguientes propiedades:*

(i) *Por lo menos una de las dos medidas, ν^+ y ν^-, es finita.*
(ii) $\nu = \nu^+ - \nu^-$
(iii) *Existen dos conjuntos $A, B \in \Im$ tales que $A \cap B = \emptyset$, $\mathbb{F} = A \cup B$ y $\nu^-(A) = \nu^+(B) = 0$.*

Además, las medidas ν^+ y ν^-, con estas propiedades, son únicas.

Demostración

Sea (A, B) una descomposición de Hahn para ν.

Para cada $E \in \Im$ definamos:

$\nu^+(E) = \nu(E \cap A)$,

$\nu^-(E) = -\nu(E \cap B)$.

ν^+ y ν^- son entonces medidas sobre $(\mathbb{F}, \Im)$ tales que:

$\nu^-(A) = \nu^+(B) = 0$ y $\nu(E) = \nu^+(E) - \nu^-(E)$ para cualquier $E \in \Im$.

Como ν toma a lo más uno de los valores $+\infty$ y $-\infty$, por lo menos una de las dos medidas, ν^+ y ν^-, es finita.

Para la unicidad, supongamos que $\left(\nu_1^+, \nu_1^-\right)$ y $\left(\nu_2^+, \nu_2^-\right)$ son dos parejas de medidas sobre $(\mathbb{F}, \Im)$ que satisfacen las propiedades 1, 2 y 3, y sean $A_1, B_1, A_2, B_2 \in \Im$ tales que:

$A_1 \cap B_1 = \emptyset$, $\mathbb{F} = A_1 \cup B_1$ y $\nu_1^-(A_1) = \nu_1^+(B_1) = 0$,

$A_2 \cap B_2 = \emptyset$, $\mathbb{F} = A_2 \cup B_2$ y $\nu_2^-(A_2) = \nu_2^+(B_2) = 0$.

Se tiene entonces:

$\nu(A_1 \cap B_2) = \nu_1^+(A_1 \cap B_2) - \nu_1^-(A_1 \cap B_2) = \nu_1^+(A_1 \cap B_2) \geq 0$,

$\nu(A_1 \cap B_2) = \nu_2^+(A_1 \cap B_2) - \nu_2^-(A_1 \cap B_2) = -\nu_2^-(A_1 \cap B_2) \leq 0$,

$\nu(A_2 \cap B_1) = \nu_1^+(A_2 \cap B_1) - \nu_1^-(A_2 \cap B_2) = -\nu_1^-(A_2 \cap B_2) \leq 0$,

$\nu(A_2 \cap B_1) = \nu_2^+(A_2 \cap B_1) - \nu_2^-(A_2 \cap B_1) = \nu_2^+(A_2 \cap B_1) \geq 0$.

Así que, $\nu(A_1 \cap B_2) = \nu(A_2 \cap B_1) = 0$.

Por lo tanto:

$\nu_1^+(A_1 \cap B_2) = \nu_1^-(A_1 \cap B_2) = 0$,

$\nu_2^+(A_1 \cap B_2) = \nu_2^-(A_1 \cap B_2) = 0$,

$\nu_1^+(A_2 \cap B_1) = \nu_1^-(A_2 \cap B_1) = 0$,

$\nu_2^+(A_2 \cap B_1) = \nu_2^-(A_2 \cap B_1) = 0$.

Entonces:

Si $C \in \Im$ y $C \subset A_1$, se tiene:

$\nu(C \cap A_2) = \nu_1^+(C \cap A_2) = \nu_1^+(C \cap A_2) + \nu_1^+(C \cap B_2) = \nu_1^+(C)$,

$\nu(C \cap A_2) = \nu_2^+(C \cap A_2) = \nu_2^+(C \cap A_2) + \nu_2^+(C \cap B_2) = \nu_2^+(C)$.

Por lo tanto:

$\nu_1^+(C) = \nu_2^+(C)$.

Así que ν_1^+ y ν_2^+ son iguales sobre A_1.

Además:

$\nu_2^+(B_1) = \nu_2^+(B_1 \cap A_2) + \nu_2^+(B_1 \cap B_2) = 0.$

Así que, ν_1^+ y ν_2^+ son iguales sobre B_1.

Por lo tanto, ν_1^+ y ν_2^+ son iguales sobre $\mathbb{F}$, es decir, $\nu_1^+ = \nu_2^+$.

Si $D \in \Im$ y $D \subset B_1$, se tiene:

$\nu(D \cap B_2) = -\nu_1^-(D \cap B_2) = -\nu_1^-(D \cap B_2) - \nu_1^-(D \cap A_2) = -\nu_1^-(D),$

$\nu(D \cap B_2) = -\nu_2^-(D \cap B_2) = -\nu_2^-(D \cap B_2) - \nu_2^-(D \cap A_2) = -\nu_2^-(D).$

Por lo tanto:

$\nu_1^-(D) = \nu_2^-(D).$

Así que ν_1^- y ν_2^- son iguales sobre B_1.

Además:

$\nu_2^-(A_1) = \nu_2^-(A_1 \cap A_2) + \nu_2^-(A_1 \cap B_2) = 0.$

Así que, ν_1^- y ν_2^- son iguales sobre A_1.

Por lo tanto, ν_1^- y ν_2^- son iguales sobre $\mathbb{F}$, es decir, $\nu_1^- = \nu_2^-$.

■

Corolario 5.11. *Sea $\nu : (\mathbb{F}, \Im) \mapsto \mathbb{R} \cup \{-\infty, \infty\}$ una medida con signo, Entonces:*

(i) *Para cualquier sucesión creciente $(A_n)_{n\in\mathbb{N}}$ de elementos de $\Im$, se tiene:*
$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightsquigarrow \infty} \nu(A_n)$.

(ii) *Para cualquier sucesión decreciente $(A_n)_{n\in\mathbb{N}}$ de elementos de $\Im$, tales que $\nu(A_1) < \infty$, se tiene:*
$\nu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \nu(A_n)$.

Demostración

Sean ν^+ y ν^- dos medidas sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ tales que:

1. Por lo menos una de las dos medidas, ν^+ y ν^-, es finita.

2. $\nu = \nu^+ - \nu^-$

3. Existen dos conjuntos $A, B \in \Im$ tales que $A \cap B = \emptyset$, $\mathbb{F} = A \cup B$ y $\nu^-(A) = \nu^+(B) = 0$.

i) Se tiene:

$\nu^+\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \nu^+(A_n)$

$\nu^-\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \nu^-(A_n)$

Como por lo menos una de las dos medidas, ν^+ y ν^-, es finita, se tiene que $\nu^+\left(\bigcup_{n=1}^{\infty} A_n\right)$ y $\nu^-\left(\bigcup_{n=1}^{\infty} A_n\right)$ no pueden ser ambas ∞. Así que:

$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) = \nu^+\left(\bigcup_{n=1}^{\infty} A_n\right) - \nu^-\left(\bigcup_{n=1}^{\infty} A_n\right) = \text{lím}_{n\rightsquigarrow\infty}\left(\nu^+(A_n) - \nu^-(A_n)\right) = \text{lím}_{n\rightsquigarrow\infty}\nu(A_n)$

ii) Como por lo menos una de las dos medidas, ν^+ y ν^-, es finita, se tiene $\nu^+(A_1) < \infty$ y $\nu^-(A_1) < \infty$. Así que:

$\nu^+\left(\bigcap_{n=1}^{\infty} A_n\right) = \text{lím}_{n\to\infty}\nu^+(A_n) < \infty.$

$\nu^-\left(\bigcap_{n=1}^{\infty} A_n\right) = \text{lím}_{n\to\infty}\nu^-(A_n) < \infty.$

Por lo tanto:

$\nu\left(\bigcap_{n=1}^{\infty} A_n\right) = \nu^+\left(\bigcap_{n=1}^{\infty} A_n\right) - \nu^-\left(\bigcap_{n=1}^{\infty} A_n\right) = \text{lím}_{n\rightsquigarrow\infty}\left(\nu^+(A_n) - \nu^-(A_n)\right) = \text{lím}_{n\rightsquigarrow\infty}\nu(A_n).$

■

Los siguientes resultados se demuestran de forma idéntica a la realizada en el caso de las medidas.

PROPOSICIÓN 5.14. *Sea $\mathbb{F}$ un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de $\mathbb{F}$ y ν_1 y ν_2 dos medidas con signo definidas sobre $\sigma(\mathcal{A})$, tales que $\nu_1(A) = \nu_2(A)$ para cualquier $A \in \mathcal{A}$. Supongamos además que existe una sucesión creciente $(F_n)_{n\in\mathbb{N}}$ de elementos de $\sigma(\mathcal{A})$ tal que $\bigcup_{n=1}^{\infty} F_n = \mathbb{F}$ y, para cualquier $n \in \mathbb{N}$, $\nu_1(F_n) = \nu_2(F_n) < \infty$ y $\nu_1(A \cap F_n) = \nu_2(A \cap F_n)$ para cualquier $A \in \mathcal{A}$. Entonces $\nu_1(A) = \nu_2(A)$ para cualquier $A \in \sigma(\mathcal{A})$.*

PROPOSICIÓN 5.15. *Sea $\mathbb{F}$ un conjunto, $\mathcal{P}$ un π-sistema de subconjuntos de $\mathbb{F}$ y ν_1 y ν_2 dos medidas con signo definidas sobre $\sigma(\mathcal{P})$, tales que $\nu_1(A) = \nu_2(A)$ para cualquier $A \in \mathcal{P}$. Supongamos además que existe una sucesión creciente $(F_n)_{n\in\mathbb{N}}$ de elementos de $\sigma(\mathcal{P})$ tal que $\bigcup_{n=1}^{\infty} F_n = \mathbb{F}$ y, para cualquier $n \in \mathbb{N}$, $\nu_1(F_n) = \nu_2(F_n) < \infty$ y $\nu_1(A \cap F_n) = \nu_2(A \cap F_n)$ para cualquier $A \in \mathcal{P}$. Entonces $\nu_1(A) = \nu_2(A)$ para cualquier $A \in \sigma(\mathcal{P})$.*

CAPÍTULO 6

MEDIDAS EN $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$

La definición de la integral de Riemann-Stieltjes de una función f con respecto a una función g, ambas definidas en un intervalo $[a, b]$, sugiere que se podría obtener una medida, sobre los borelianos de $[a, b]$, partiendo de que la medida de un intervalo de extremos x_1 y x_2, ambos en el intervalo $[a, b]$, puede definirse como la diferencia $g(x_2) - g(x_1)$. En general, esa diferencia podría ser positiva o negativa, de manera que estaríamos generando una medida con signo a partir de la función g. Como una medida con signo se puede expresar como la diferencia de dos medidas, podemos analizar primero bajo que condiciones una función $h : [a, b] \to \mathbb{R}$ genera una medida partiendo, como se indicó antes, de que la medida de un intervalo es la diferencia de los valores de h en sus extremos.

En primer lugar, es evidente que, para que la medida de los intervalos así definida sea no negativa, se requiere que h sea una función no decreciente. De manera que, dada una función g, para que se pueda generar una medida partiendo de que la medida de un intervalo es la diferencia de los valores de g en sus extremos, se requiere que g sea de variación acotada.

Tomemos entonces una función $h : [a, b] \to \mathbb{R}$ no decreciente y supongamos que podemos definir una medida μ_h, sobre los borelianos de $[a, b]$, tal que la medida de un intervalo es la diferencia de los valores de h en sus extremos.

Si $\mu_h((c, d)) = h(d) - h(c)$ para cualquier intervalo (c, d) contenido en (a, b), entonces, dado uno de esos intervalos, si $((c_n, d_n))_{n\in\mathbb{N}}$ es una suceción creciente de intervalos cuya unión es el intervalo (c, d), se tiene $h(d) - h(c) = \mu_h((c, d)) = \text{lím}_{n\rightsquigarrow\infty} \mu_h((c_n, d_n)) = \text{lím}_{n\rightsquigarrow\infty} (h(d_n) - h(c_n))$, así que h tiene que ser continua por la derecha en c y continua por la izquierda en d. Siendo esto válido para cualquier intervalo (c, d) contenido en $[a, b]$, h tiene que ser continua en el intervalo (a, b). En este caso, la medida de un intervalo, de cualquier tipo, de extremos c y d, con $c, d \in (a, b)$, es igual a $h(d) - h(c)$.

De la misma manera:

Si $\mu_h([c, d]) = h(d) - h(c)$ para cualquier intervalo $[c, d]$ contenido en (a, b), entonces h tiene que ser continua en el intervalo (a, b) y la medida de un intervalo, de cualquier tipo, de extremos c y d, con $c, d \in (a, b)$, es igual a $h(d) - h(c)$.

Si $\mu_h([c,d)) = h(d) - h(c)$ para cualquier intervalo $[c,d)$ contenido en (a,b), entonces h tiene que ser continua por la izquierda en el intervalo (a,b).

Si $\mu_h((c,d]) = h(d) - h(c)$ para cualquier intervalo $(c,d]$ contenido en (a,b), entonces h tiene que ser continua por la derecha en el intervalo (a,b).

Vamos a demostrar que, en cualquiera de estos cuatro casos, se puede generar la medida μ_h. Basta con demostrarlo para los dos últimos casos ya que cualquiera de ellos contiene, como casos particulares, a los dos primeros.

El método para generar la medida μ_h será el que expusimos en la sección 5.3 del capítulo anterior. Es decir, definiendo primero una quasi medida sobre el álgebra generada por los intervalos de la forma $[c,d)$ para el caso de una función no decreciente continua por la izquierda o por los intervalos de la forma $(c,d]$ para el caso de una función no decreciente continua por la derecha.

Vamos a tratar un caso más general, el de una función $F : \mathbb{R} \mapsto \mathbb{R}$ no decreciente, para la cual demostraremos que existe una única medida μ_F, definida sobre los conjuntos borelianos de $\mathbb{R}$, tal que, para cualquier pareja de números reales, a y b, tales que $a < b$, se tiene:

$$\mu_F((a,b]) = F(b+) - F(a+),$$

o equivalentemente:

$$\mu_F([a,b)) = F(b-) - F(a-).$$

Una medida con esta propiedad es finita sobre los intervalos acotados.

Demostraremos entonces el resultado inverso, es decir, que dada cualquier medida μ sobre los conjuntos borelianos de $\mathbb{R}$, tal que los intervalos acotados tienen medida finita, existe una función no decreciente $F_\mu : \mathbb{R} \mapsto \mathbb{R}$ tal que, para cualquier pareja de números reales, a y b, tales que $a < b$, se tiene:

$$\mu((a,b]) = F_\mu(b+) - F_\mu(a+),$$

$$\mu([a,b)) = F_\mu(b-) - F_\mu(a-).$$

Una función no decreciente F_μ con esta propiedad no es única ya que, por ejemplo, si $c \in \mathbb{R}$, la función $F_\mu + c$ es no decreciente y cumple con la misma propiedad. Sin embargo, lo que se puede probar es que si F_1 y F_2 son dos funciones no decrecientes que satisfacen esta propiedad, entonces las funciones $x \to (F_1 - F_2)(x+)$ y $x \to (F_1 - F_2)(x-)$, definidas sobre $\mathbb{R}$, son constantes.

Si nos restringimos a las funciones no decrecientes y continuas por la derecha o a las no decrecientes y continuas por la izquierda, entonces, dadas dos funciones F_1 y F_2 que satisfacen dicha propiedad, existe $c \in \mathbb{R}$ tal que $F_2 - F_1 = c$.

Por lo tanto, si identificamos a las funciones no decrecientes continuas por la derecha (resp. continuas por la izquierda) cuya diferencia es constante, podemos afirmar que hay una correspondencia uno a uno entre las funciones no decrecientes continuas por la derecha (resp. continuas por la izquierda) y las medidas definidas sobre los conjuntos borelianos de $\mathbb{R}$ tal que los intervalos acotados tienen medida finita. De manera más específica, si $F_1 : \mathbb{R} \mapsto \mathbb{R}$ y $F_2 : \mathbb{R} \mapsto \mathbb{R}$ son dos funciones no decrecientes continuas por la derecha (resp. continuas por la izquierda), diremos que son equivalentes si existe $c \in \mathbb{R}$ tal que $F_2 - F_1 = c$. Denotemos por $\mathcal{M}$ al conjunto de medidas, definidas sobre los conjuntos borelianos de $\mathbb{R}$, tal que los intervalos acotados tienen medida finita, y por $\mathcal{C}^+$ (resp. $\mathcal{C}^-$) al conjunto de clases de equivalencia en las cuales queda partido el conjunto de funciones no decrecientes continuas por la derecha (resp. continuas por la izquierda) bajo la relación de equivalencia así definida. Entonces, existe una función biyectiva $H : \mathcal{C}^+ \to \mathcal{M}$ (resp. $H : \mathcal{C}^- \to \mathcal{M}$) tal que si $\mu_X = H(X)$, entonces, para cualquier pareja de números reales, a y b, tales que $a < b$, se tiene:

$$\mu_X((a,b]) = F(b) - F(a) \text{ (resp. } \mu_X([a,b)) = F(b) - F(a))$$

para cualquier función $F \in X$.

En otras palabras, toda función no decreciente continua por la derecha (resp. continua por la izquierda) representa una única medida, definida sobre los conjuntos borelianos de $\mathbb{R}$, tal que los intervalos acotados tienen medida finita. Inversamente, toda medida definida sobre los conjuntos borelianos de $\mathbb{R}$ tal que los intervalos acotados tienen medida finita representa una única (módulo la relación de equivalencia definida) función no decreciente continua por la derecha (resp. continua por la izquierda).

6.1. Medidas y funciones no decrecientes

Sea $F : \mathbb{R} \mapsto \mathbb{R}$ una función no decreciente y continua por la derecha y definamos:

$F(\infty) = \text{lím}_{x \rightsquigarrow \infty} F(x)$,

$F(-\infty) = \text{lím}_{x \to -\infty} F(x)$.

Si $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ y $a < b$, definamos $(a, b|$ de la siguiente manera:

$$(a,b| = \begin{cases} (a,b] & \text{si } b \in \mathbb{R} \\ (a,b) & \text{si } b = \infty \end{cases}$$

Sea $\mathcal{I}$ la familia de los intervalos de este tipo, agregando al vacío como parte de la familia.

Definamos $\nu_F(\emptyset) = 0$ y, para cada intervalo $I = (a,b| \in \mathcal{I}$, $\nu_F(I) = F(b) - F(a)$.

Se podría tener $F(\infty) = \infty$ y $F(-\infty) = -\infty$, lo cual no sería problema para la definición de $\nu_F((-\infty, \infty))$ pues, con las convenciones que hemos hecho, se tendría $\mu_F((-\infty, \infty)) = \infty - (-\infty) = \infty + \infty = \infty$.

El resultado central que se requiere demostrar para poder extender ν_F a la σ-álgebra de los conjuntos borelianos en $\mathbb{R}$ es el siguiente:

Sea $(a, b| \in \mathcal{I}$ y $(a_1, b_1|, (a_2, b_2|, \ldots$ una colección infinita de intervalos en $\mathcal{I}$ tales que $(a, b| \subset \bigcup_{k=1}^{\infty} (a_k, b_k|$. Entonces:

$$F(b) - F(a) \leq \sum_{k=1}^{\infty} [F(b_k) - F(a_k)].$$

Lema 6.1. *Sea $I = (a, b| \in \mathcal{I}$ y $\left(a^{(1)}, b^{(1)}\right|, \left(a^{(2)}, b^{(2)}\right|, \ldots, \left(a^{(m)}, b^{(m)}\right|$, una colección finita de intervalos en $\mathcal{I}$, ajenos por parejas, tal que $I = \bigcup_{j=1}^{m} \left(a^{(j)}, b^{(j)}\right|$, entonces:*

$\nu_F(I) = \sum_{j=1}^{m} \nu_F\left(\left(a^{(j)}, b^{(j)}\right|\right)$.

Demostración

Como los intervalos $\left(a^{(1)}, b^{(1)}\right|, \left(a^{(2)}, b^{(2)}\right|, \ldots, \left(a^{(m)}, b^{(m)}\right|$ son ajenos por parejas y su unión es $(a, b|$, podemos ordenarlos para obtener una colección de intervalos ajenos por parejas, $\left(x^{(1)}, y^{(1)}\right|, \left(x^{(2)}, y^{(2)}\right|, \ldots, \left(x^{(m)}, y^{(m)}\right|$, de tal forma que:

$a = x^{(1)} < y^{(1)} = x^{(2)} < y^{(2)} = x^{(2)} < \cdots = x^{(m)} < y^{(m)} = b$

Así que:

$\sum_{j=1}^{m} \nu_F\left(\left(a^{(j)}, b^{(j)}\right|\right) = \sum_{j=1}^{m} \nu_F\left(\left(y^{(j)}, z^{(j)}\right|\right) = \sum_{j=1}^{m} \left[F\left(y^{(j)}\right) - F\left(x^{(j)}\right)\right]$

$= F(b) - F(a) = \mu_F(I)$

■

Lema 6.2. *Sea $I = (a, b| \in \mathcal{I}$ y $I^{(1)} = \left(a^{(1)}, b^{(1)}\right|, \ldots, I^{(m)} = \left(a^{(m)}, b^{(m)}\right|$ una colección finita de intervalos en $\mathcal{I}$ tales que $I \subset \bigcup_{j=1}^{m} I^{(j)}$, entonces:*

$\nu_F(I) \leq \sum_{j=1}^{m} \nu_F\left(I^{(j)}\right)$.

Demostración

Los puntos $a, b, a^{(1)}, b^{(1)}, \ldots, a^{(m)}, b^{(m)}$ constituyen una partición de un intervalo $(c, d|$ que contiene al intervalo $(a, b|$.

Esta partición parte cada intervalo $I^{(j)}$, con $j \in \{1, 2, \ldots, m\}$, en subintervalos ajenos por parejas, $I_1^{(j)}, \ldots, I_{n_j}^{(j)}$. Así que:

$\nu_F\left(I^{(j)}\right) = \sum_{k=1}^{n_j} \nu_F\left(I_k^{(j)}\right)$.

La partición definida antes también parte el intervalo I en subintervalos ajenos por parejas, $I_1, \ldots, I_n$. Así que:

$\nu_F(I) = \sum_{k=1}^{n} \nu_F(I_k)$.

Por otra parte, como $I \subset \bigcup_{j=1}^{m} I^{(j)}$, cada intervalo I_k, con $k \in \{1, 2, \ldots, n\}$, coincide con un intervalo $I_{k'}^{(j)}$ para alguna $j \in \{1, 2, \ldots, m\}$ y alguna $k' \in \{1, 2, \ldots, n_j\}$, por lo tanto:

$\nu_F(I) = \sum_{k=1}^{n} \nu_F(I_k) \leq \sum_{j=1}^{m} \sum_{k=1}^{n_j} \nu_F\left(I_k^{(j)}\right) = \sum_{j=1}^{m} \nu_F\left(I^{(j)}\right)$.

■

Lema 6.3. *Sean $I_1, \ldots, I_k$ y $I^{(1)}, \ldots, I^{(m)}$ dos colecciones finitas de intervalos en $\mathcal{I}$ tales que $I_1, \ldots, I_k$ son ajenos por parejas, $I^{(1)}, \ldots, I^{(m)}$ son ajenos por parejas y $\bigcup_{i=1}^{k} I_i = \bigcup_{j=1}^{m} I^{(j)}$, entonces:*

$\sum_{i=1}^{k} \nu_F(I_i) = \sum_{j=1}^{m} \nu_F\left(I^{(j)}\right)$.

Demostración

Para cada $i \in \{1, \ldots, k\}$ y $j \in \{1, \ldots, m\}$, definamos $I_i^{(j)} = I_i \cap I^{(j)}$. Entonces, como $\bigcup_{i=1}^{k} I_i = \bigcup_{j=1}^{m} I^{(j)}$, se tiene $I_i = \bigcup_{j=1}^{m} I_i^{(j)}$ y $I^{(j)} = \bigcup_{i=1}^{k} I_i^{(j)}$, así que:

$\nu_F(I_i) = \sum_{j=1}^{m} \nu_F\left(I_i^{(j)}\right)$,

$\nu_F\left(I^{(j)}\right) = \sum_{i=1}^{k} \nu_F\left(I_i^{(j)}\right)$.

Por lo tanto:

$\sum_{i=1}^{k} \nu_F(I_i) = \sum_{i=1}^{k} \sum_{j=1}^{m} \nu_F\left(I_i^{(j)}\right)$

$= \sum_{j=1}^{m} \sum_{i=1}^{k} \nu_F\left(I_i^{(j)}\right) = \sum_{j=1}^{m} \nu_F\left(I^{(j)}\right)$.

■

Teorema 6.1. *Sea $(a, b| \in \mathcal{I}$ y $(a_1, b_1|, (a_2, b_2|, \ldots$ una colección infinita de intervalos en $\mathcal{I}$ tales que $(a, b| \subset \bigcup_{k=1}^{\infty} (a_k, b_k|$. Entonces:*

$F(b) - F(a) \leq \sum_{k=1}^{\infty} [F(b_k) - F(a_k)]$.

Demostración

Tomemos $\varepsilon > 0$ y $\delta > 0$, arbitrarios.

Como F es continua por la derecha, para cada cada $k \in \mathbb{N}$, existe $\delta_k > 0$ tal que:

$F(d_k) - F(b_k) < \frac{\varepsilon}{2^k}$,

donde:

$$d_k = \begin{cases} b_k + \delta_k & \text{si } b_k \in \mathbb{R} \\ b_k & \text{si } b_k = \infty \end{cases}$$

Definamos:

$$c_\delta = \begin{cases} a + \delta & \text{si } a \in \mathbb{R} \\ -\frac{1}{\delta} & \text{si } a = -\infty \end{cases}$$

$$d_\delta = \begin{cases} b & \text{si } b \in \mathbb{R} \\ \frac{1}{\delta} & \text{si } b = \infty \end{cases}$$

Entonces:

$\text{lím}_{\delta \to 0} \left[F(d_\delta) - F(c_\delta) \right] = F(b) - F(a)$.

Además:

$(c_\delta, d_\delta| \subset [c_\delta, d_\delta] \subset (a, b| \subset \bigcup_{k=1}^{\infty} (a_k, b_k| \subset \bigcup_{k=1}^{\infty} (a_k, d_k)$.

Así que, por el teorema de Heine-Borel, existe una colección finita, $(a_{k_1}, d_{k_1}), \ldots, (a_{k_m}, d_{k_m})$, tal que:

$[c_\delta, d_\delta] \subset \bigcup_{j=1}^{m} \left(a_{k_j}, d_{k_j} \right)$

Por lo tanto:

$(c_\delta, d_\delta| \subset [c_\delta, d_\delta] \subset \bigcup_{j=1}^{m} \left(a_{k_j}, d_{k_j} \right) \subset \bigcup_{j=1}^{m} \left(a_{k_j}, d_{k_j} \right|$.

Así que:

$F(d_\delta) - F(c_\delta) \leq \sum_{j=1}^{m} \left[F\left(d_{k_j}\right) - F\left(a_{k_j}\right) \right] \leq \sum_{k=1}^{\infty} \left[F(d_k) - F(a_k) \right]$

$\leq \sum_{k=1}^{\infty} \left[F(b_k) - F(a_k) \right] + \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \sum_{k=1}^{\infty} \left[F(b_k) - F(a_k) \right] + \varepsilon$.

Y, como $\varepsilon > 0$ es arbitraria:

$F(d_\delta) - F(c_\delta) \leq \sum_{k=1}^{\infty} \left[F(b_k) - F(a_k) \right]$.

Finalmente, tomando límites cuando $\delta \to 0$, se obtiene:

$F(b) - F(a) \leq \sum_{k=1}^{\infty} \left[F(b_k) - F(a_k) \right]$.

■

Teorema 6.2. *Si $F : \mathbb{R} \mapsto \mathbb{R}$ es una función no decreciente y continua por la derecha, existe una única medida μ_F, definida sobre los conjuntos borelianos de $\mathbb{R}$, tal que:*

$\mu_F((a, b]) = F(b) - F(a)$,

para cualquier pareja de números reales, a y b, tales que $a < b$.

Además, $\mu_F(\{x\}) = F(x) - F(x-)$ *para cualquier* $x \in \mathbb{R}$.

Demostración

Sea $\mathcal{A}$ la familia formada por los conjuntos de la forma $\bigcup_{j=1}^{n} I_j$ donde $n \in \mathbb{N}$ y $I_1, \dots, I_n$ son intervalos en $\mathcal{I}$, ajenos por parejas.

Para cada $A = \bigcup_{j=1}^{n} I_j \in \mathcal{A}$, definamos $\mu_F(A) = \sum_{j=1}^{n} \nu_F(I_j)$.

Por los lemas anteriores, μ_F está bien definida.

Obviamente $\nu_F(I) = \mu_F(I)$ para cualquier $I \in \mathcal{I}$, $\mathcal{A}$ es un álgebra de subconjuntos de $\mathbb{R}$ y la función $\mu_F : \mathcal{A} \mapsto \mathbb{R}$ es no negativa y finitamente aditiva.

Sea $A_1, A_2, \dots$ una colección infinita numerable de elementos de $\mathcal{A}$, ajenos por parejas, no vacíos y tales que $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Como para cada $i \in \mathbb{N}$, $A_i \in \mathcal{A}$, A_i es una unión finita de intervalos en $\mathcal{I}$ ajenos por parejas. Además, como $A \in \mathcal{A}$, A también es una unión finita de intervalos en $\mathcal{I}$ ajenos por parejas.

Sean $A = \bigcup_{j=1}^{m} I^{(j)}$ y, para cada $i \in \mathbb{N}$, $A_i = \bigcup_{k=1}^{m_i} I_{(i,k)}$. Entonces:

$\bigcup_{j=1}^{m} I^{(j)} = A = \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \bigcup_{k=1}^{m_i} I_{(i,k)}$

Tenemos dos colecciones de intervalos, por un lado la familia $\left\{ I_{(i,k)} : i \in \mathbb{N}, k \in \{1, \dots, m_i\} \right\}$ y por el otro la familia $\left\{ I^{(j)} : j \in \{1, \dots, m\} \right\}$. Tanto los intervalos de la primera familia como los de la segunda son ajenos por parejas y A es igual tanto a la unión de los intervalos de la primera familia como de la segunda. Por otra parte, una pareja de intervalos, uno de la primera familia y otro de la segunda, podrían no ser ajenos.

La idea ahora es partir cada intervalo $I^{(j)}$ en intervalos ajenos por parejas, utilizando los intervalos $I_{(i,k)}$. Para esto, definamos, para cada $j \in \{1, \dots, m\}$, $i \in \mathbb{N}$ y $k \in \{1, \dots, m_i\}$:

$I_{(i,k)}^{(j)} = I_{(i,k)} \cap I^{(j)}$

Entonces, los intervalos de la familia $\left\{ I_{(i,k)}^{(j)} : j \in \{1, \dots, m\}, i \in \mathbb{N}, k \in \{1, \dots, m_i\} \right\}$ son ajenos por parejas y:

$I_{(i,k)} = \bigcup_{j=1}^{m} I_{(i,k)}^{(j)}$ para cualesquiera $i \in \mathbb{N}$ y $k \in \{1, \dots, m_i\}$.

$I^{(j)} = \bigcup_{i=1}^{\infty} \bigcup_{k=1}^{m_i} I_{(i,k)}^{(j)}$ para cualquier $j \in \{1, \dots, m\}$.

Además, por el teorema 6.1, se tiene:

$\mu_F\left(I^{(j)}\right) \leq \sum_{i=1}^{\infty} \sum_{k=1}^{m_i} \mu_F\left(I_{(i,k)}^{(j)}\right)$ para cualquier $j \in \{1, \dots, m\}$.

Así que:

$\mu_F(A) = \sum_{j=1}^{m} \mu_F\left(I^{(j)}\right) \leq \sum_{j=1}^{m} \sum_{i=1}^{\infty} \sum_{k=1}^{m_i} \mu_F\left(I_{(i,k)}^{(j)}\right)$

$= \sum_{i=1}^{\infty} \sum_{k=1}^{m_i} \sum_{j=1}^{m} \mu_F\left(I_{(i,k)}^{(j)}\right) = \sum_{i=1}^{\infty} \sum_{k=1}^{m_i} \mu_F\left(I_{(i,k)}\right)$

$= \sum_{i=1}^{\infty} \mu_F(A_i)$

Además, como μ_F es finitamente aditiva y $A \supset \bigcup_{i=1}^{n} A_i$ para cualquier $n \in \mathbb{N}$, se tiene $\mu_F(A) \geq \sum_{i=1}^{n} \mu_F(A_i)$ para cualquier $n \in \mathbb{N}$, así que:

$\mu_F(A) \geq \sum_{i=1}^{\infty} \mu_F(A_i)$.

Por lo tanto, $\mu_F(A) = \sum_{i=1}^{\infty} \mu_F(A_i)$, así que μ_F es σ-aditiva y entonces puede ser extendida a una medida μ_F definida sobre la σ-álgebra generada por $\mathcal{A}$, es decir, los borelianos de $\mathbb{R}$.

Si μ_1 y μ_2 son dos medidas sobre los borelianos de $\mathbb{R}$ tales que $\mu_1((a,b]) = \mu_2((a,b]) = F(b) - F(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$, definamos, para cada $n \in \mathbb{N}$, $F_n = (-n, n]$; entonces $\bigcup_{n=1}^{\infty} F_n = \mathbb{R}$ y $\mu_1(F_n) = \mu_2(F_n) < \infty$ para cualquier $n \in \mathbb{N}$. Así que, por el teorema de clases monótonas para π-sistemas, $\mu_1(A) = \mu_2(A)$ para cualquier conjunto boreliano de $\mathbb{R}$.

Para la última parte, sea $x \in \mathbb{R}$, entonces:

$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x\right]$.

Así que:

$\mu_F(\{x\}) = \lim_{n \rightsquigarrow \infty} \mu_F\left(\left(x - \frac{1}{n}, x\right]\right) = F(x) - \lim_{n \rightsquigarrow \infty} F\left(x - \frac{1}{n}\right) = F(x) - F(x-)$.

■

Corolario 6.1. *Si $F : \mathbb{R} \mapsto \mathbb{R}$ es una función no decreciente y continua por la izquierda, existe una única medida μ_F, definida sobre los conjuntos borelianos de $\mathbb{R}$, tal que:*

$\mu_F([a,b)) = F(b) - F(a)$,

para cualquier pareja de números reales, a y b, tales que $a < b$.

Además, $\mu_F(\{x\}) = F(x+) - F(x)$ para cualquier $x \in \mathbb{R}$.

Demostración

Definamos $F^d : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:

$F^d(x) = F(x+)$.

Entonces, F^d es no decreciente y continua por la derecha; además, $F^d(x) - F^d(x-) = F(x+) - F(x)$ para cualquier $x \in \mathbb{R}$, lo cual implica que, si a y b son números reales tales que $a < b$, entonces:

$F^d(b-) - F^d(a-) = F(a+) - F(a) - F^d(a) - F(b+) + F(b) + F^d(b)$

$= F(a+) - F(a) - F(a+) - F(b+) + F(b) + F(b+)$

$= F(b) - F(a).$

Sea μ_F, la una única medida, definida sobre los conjuntos borelianos de $\mathbb{R}$, tal que:

$\mu_F((a,b]) = F^d(b) - F^d(a),$

para cualquier pareja de números reales, a y b, tales que $a < b$.

Sean $a, b \in \mathbb{R}$ tales que $a < b$, $(a_n)_{n\in\mathbb{N}}$ una sucesión creciente que converja a a y $(b_n)_{n\in\mathbb{N}}$ una sucesión creciente, de números reales mayores que a, que converja a b; entonces:

$\mu_F([a,b)) = \text{lím}_{n\rightsquigarrow\infty} \mu_F((a_n, b_n]) = \text{lím}_{n\to\infty} F^d(b_n) - \text{lím}_{n\rightsquigarrow\infty} F^d(a_n)$

$= F^d(b-) - F^d(a-) = F(b) - F(a).$

Si μ_1 y μ_2 son dos medidas sobre los borelianos de $\mathbb{R}$ tales que $\mu_1([a,b)) = \mu_2([a,b)) = F(b) - F(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$, definamos, para cada $n \in \mathbb{N}$, $F_n = [-n, n)$; entonces $\bigcup_{n=1}^{\infty} F_n = \mathbb{R}$ y $\mu_1(F_n) = \mu_2(F_n) < \infty$ para cualquier $n \in \mathbb{N}$. Así que, por el teorema de clases monótonas para π-sistemas, $\mu_1(A) = \mu_2(A)$ para cualquier conjunto boreliano de $\mathbb{R}$.

Para la última parte, sea $x \in \mathbb{R}$, entonces:

$\{x\} = \bigcap_{n=1}^{\infty} \left[x, x + \frac{1}{n}\right).$

Así que:

$\mu_F(\{x\}) = \text{lím}_{n\rightsquigarrow\infty} \mu_F\left(\left[x, x + \frac{1}{n}\right)\right) = \text{lím}_{n\to\infty} F\left(x + \frac{1}{n}\right) - F(x) = F(x+) - F(x).$

■

Sea $F : \mathbb{R} \mapsto \mathbb{R}$ una función no decreciente, entonces:

La función $F^{(d)} : \mathbb{R} \to \mathbb{R}$ definida por $F^{(d)}(x) = F(x+)$ es no decreciente y continua por la derecha, y $F^{(d)}(x-) = F(x-)$ para cualquier $x \in \mathbb{R}$.

La función $F^{(i)} : \mathbb{R} \to \mathbb{R}$ definida por $F^{(i)}(x) = F(x-)$ es no decreciente y continua por la izquierda, y $F^{(i)}(x+) = F(x+)$ para cualquier $x \in \mathbb{R}$.

Así que, de acuerdo con los resultados anteriores, podemos generar 2 medidas sobre los conjuntos borelianos de $\mathbb{R}$, $\mu_{F^{(d)}}$ y $\mu_{F^{(i)}}$, tales que, para cualquier pareja de números reales, a y b, tales que $a < b$, se tiene:

$\mu_{F^{(d)}}((a,b]) = F^{(d)}(b) - F^{(d)}(a),$

$\mu_{F^{(i)}}([a,b)) = F^{(i)}(b) - F^{(i)}(a).$

En realidad, estas 2 medidas son una sola. En efecto, si a y b son números reales tales que $a < b$, se tiene:

$\mu_{F^i}((a,b]) = \text{lím}_{n\to\infty} \mu_{F^i}\left(\left[a + \frac{1}{n}, b + \frac{1}{n}\right)\right) = \text{lím}_{n\rightsquigarrow\infty}\left[F^{(i)}\left(b + \frac{1}{n}\right) - F^{(i)}\left(a + \frac{1}{n}\right)\right]$

$= F^{(i)}(b+) - \text{lím}_{n\rightsquigarrow\infty} F^{(i)}(a+) = F(b+) - F(a+)$

$= F^{(d)}(b) - F^{d}(a) = \mu_{F^{(d)}}((a,b])$.

Así que $\mu_{F^i} = \mu_{F^{(d)}}$.

Así que, los 2 resultados anteriores se reducen a un único resultado:

Teorema 6.3. *Si $F : \mathbb{R} \mapsto \mathbb{R}$ es una función no decreciente, existe una única medida μ_F, definida sobre los conjuntos borelianos de $\mathbb{R}$, tal que:*

(i) $\mu_F((a,b]) = F(b+) - F(a+)$,
(ii) $\mu_F([a,b)) = F(b-) - F(a-)$,
(iii) $\mu_F(\{x\}) = F(x+) - F(x-)$,
para cualquier $x \in \mathbb{R}$ y cualquier pareja de números reales, a y b, tales que $a < b$.

Definición 6.1. *Si $F : \mathbb{R} \mapsto \mathbb{R}$ es una función no decreciente, la medida μ_F del teorema anterior será llamada la medida generada por F.*

Ahora mostraremos que cualquier medida sobre los borelianos de $\mathbb{R}$, tal que los intervalos acotados tienen medida finita, se puede generar a partir de una función no decreciente ya sea continua por la derecha o continua por la izquierda.

Teorema 6.4. *Dada cualquier medida μ sobre los conjuntos borelianos de $\mathbb{R}$, tal que los intervalos acotados tienen medida finita, existe una función no decreciente y continua por la derecha $F_\mu^d : \mathbb{R} \mapsto \mathbb{R}$ tal que $\mu((a,b]) = F_\mu^d(b) - F_\mu^d(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$. Además, si F_1 y F_2 son dos funciones con esta propiedad, entonces $F_1 - F_2$ es constante.*

Demostración

Definamos la función $F_\mu^d : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:

$$F_\mu^d(x) = \begin{cases} -\mu((x,0]) & \text{si } x \in (-\infty, 0) \\ 0 & \text{si } x = 0 \\ \mu((0,x]) & \text{si } x \in (0,\infty) \end{cases}$$

F_μ^d es no decreciente y continua por la derecha y se tiene $\mu((a,b]) = F_\mu^d(b) - F_\mu^d(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$.

Si $F_1 : \mathbb{R} \mapsto \mathbb{R}$ y $F_2 : \mathbb{R} \mapsto \mathbb{R}$ son dos funciones que satisfacen el enunciado del teorema, se tiene $F_1(b) - F_1(a) = F_2(b) - F_2(a)$ para cualquier pareja de números reales a y b tales que $a < b$, lo cual implica que $(F_1 - F_2)(b) = (F_1 - F_2)(a)$; es decir, $F_1 - F_2$ es constante.

■

Corolario 6.2. *Para cualquier medida μ sobre los conjuntos borelianos de $\mathbb{R}$, tal que los intervalos acotados tienen medida finita, existe una función no decreciente y continua por la izquierda $F_\mu^i : \mathbb{R} \mapsto \mathbb{R}$ tal que $\mu([a,b)) = F_\mu^i(b) - F_\mu^i(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$. Además, si F_1 y F_2 son dos funciones con esta propiedad, entonces $F_1 - F_2$ es constante.*

Demostración

Sea F_μ^d una función no decreciente y continua por la derecha $F_\mu^d : \mathbb{R} \mapsto \mathbb{R}$ tal que $\mu((a,b]) = F_\mu^d(b) - F_\mu^d(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$.

Definamos $F_\mu^i : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:

$F_\mu^i(x) = F_\mu^d(x-)$.

Entonces, F_μ^i es no decreciente y continua por la izquierda.

Sean $a, b \in \mathbb{R}$ tales que $a < b$, $(a_n)_{n\in\mathbb{N}}$ una sucesión creciente que converja a a y $(b_n)_{n\in\mathbb{N}}$ una sucesión creciente, de números reales mayores que a, que converja a b; entonces:

$\mu([a,b)) = \lim_{n\rightsquigarrow\infty} \mu((a_n, b_n]) = \lim_{n\rightsquigarrow\infty} F_\mu^d(b_n) - \lim_{n\rightsquigarrow\infty} F_\mu^d(b_n)$

$= F_\mu^d(b-) - F_\mu^d(a-) = F_\mu^i(b) - F_\mu^i(a)$.

Si $F_1 : \mathbb{R} \mapsto \mathbb{R}$ y $F_2 : \mathbb{R} \mapsto \mathbb{R}$ son dos funciones que satisfacen el enunciado del teorema, se tiene $F_1(b) - F_1(a) = F_2(b) - F_2(a)$ para cualquier pareja de números reales a y b tales que $a < b$, lo cual implica que $(F_1 - F_2)(b) = (F_1 - F_2)(a)$; es decir, $F_1 - F_2$ es constante.

■

Como antes, los 2 resultados anteriores se reducen a un único resultado:

Teorema 6.5. *Dada cualquier medida μ sobre los conjuntos borelianos de $\mathbb{R}$, tal que los intervalos acotados tienen medida finita, existe una función no decreciente $F_\mu : \mathbb{R} \mapsto \mathbb{R}$ tal que la medida generada por F_μ es μ. Es decir, se tiene:*

(i) $\mu((a,b]) = F_\mu(b+) - F_\mu(a+)$,
(ii) $\mu([a,b)) = F_\mu(b-) - F_\mu(a-)$,
(iii) $\mu(\{x\}) = F_\mu(x+) - F_\mu(x-)$,

para cualquier $x \in \mathbb{R}$ y cualquier pareja de números reales, a y b, tales que $a < b$.

Además, si F_1 y F_2 son dos funciones no decrecientes con estas propiedades, entonces las funciones $x \to (F_1 - F_2)(x+)$ y $x \to (F_1 - F_2)(x-)$, definidas sobre $\mathbb{R}$, son constantes.

Corolario 6.3. *Si μ es una medida sobre los conjuntos borelianos de $\mathbb{R}$, tal que los intervalos acotados tienen medida finita, el conjunto $\{x \in \mathbb{R} : \mu(\{x\}) > 0\}$ es a lo más infinito numerable.*

Se tiene el siguiente resultado que es más general:

Proposición 6.1. *Si μ es una medida σ-finita sobre los conjuntos borelianos de $\mathbb{R}$, el conjunto $\{x \in \mathbb{R} : \mu(\{x\}) > 0\}$ es a lo más infinito numerable.*

Demostración

Sea $E_1, E_2, \ldots$ una colección infinita numerable de conjuntos borelianos de $\mathbb{R}$ tal que $\bigcup_{n=1}^{\infty} E_n = \mathbb{R}$ y $\mu(E_n) < \infty$ para cualquier $n \in \mathbb{N}$.

Para cada pareja $n, m \in \mathbb{N}$, definamos:

$E_n^{(m)} = \left\{x \in E_n : \mu(\{x\}) > \frac{1}{m}\right\}$.

Como $\mu(E_n) < \infty$, $E_n^{(m)}$ es un conjunto finito para cualquier $m \in \mathbb{N}$. Además:

$\{x \in E_n : \mu(\{x\}) > 0\} = \bigcup_{m=1}^{\infty} E_n^{(m)}$.

Así que el conjunto $\{x \in E_n : \mu(\{x\}) > 0\}$ es a lo más infinito numerable.

Finalmente, $\{x \in \mathbb{R} : \mu(\{x\}) > 0\} = \bigcup_{n=1}^{\infty} \{x \in E_n : \mu(\{x\}) > 0\}$. Así que el conjunto $\{x \in \mathbb{R} : \mu(\{x\}) > 0\}$ es a lo más infinito numerable.

■

6.2. Medidas y funciones no decrecientes que crecen únicamente mediante saltos

Como vimos en el capítulo 2, toda función no decreciente se puede expresar como la suma de dos funciones no decrecientes, una que crece únicamente mediante saltos y la otra continua. Por otra parte, una medida sobre los conjuntos borelianos de $\mathbb{R}$, tal que los intervalos acotados tienen medida finita, podemos generarla ya sea mediante una función no decreciente y continua por la derecha, o bien mediante una no decreciente y continua por la izquierda. Se hizo costumbre elegir la función continua por la derecha para representar a la medida, pero bien podría elegirse la que es continua por la izquierda.

Tomemos una función $F : \mathbb{R} \mapsto \mathbb{R}$, no decreciente y continua por la derecha. F se puede expresar entonces como la suma de una función F^c, continua, y una función F^d, no decreciente, continua por la derecha, que crece únicamente mediante saltos y tal que $F^d(x) - F^d(x-) =$

$F(x) - F(x-)$ para cualquier $x \in \mathbb{R}$. Además, tenemos la definición explícita de la función F^d:

$$F^d(x) = \begin{cases} F(0-) - \sum_{\{y \in D: x < y < 0\}} [F(y) - F(y-)] & \text{si } x \in (-\infty, 0) \\ F(0) + \sum_{\{y \in D: 0 < y \leq x\}} [F(y) - F(y-)] & \text{si } x \in [0, \infty) \end{cases}$$

Donde D es el conjunto de puntos donde F es discontinua, el cual es a lo más infinito numerable y lo supondremos no vacío.

Si F^c es constante, la medida que genera es nula sobre los borelianos. En este caso, la medida generada por F es la generada por F^d. Si denotamos por μ_{F^d} a la medida generada por F^d, recordemos que se tienen las siguientes relaciones:

$F^d(z) - F^d(x) = \sum_{\{y \in D: x < y \leq z\}} \left[F^d(y) - F^d(y-)\right]$ para cualquier parera (x, z) de números reales tales que $x < z$.

$\mu_{F^d}((a, b]) = F^d(b) - F^d(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$.

$\mu_{F^d}(\{x\}) = F^d(x) - F^d(x-)$ para cualquier $x \in \mathbb{R}$; en particular, $\mu_{F^d}(\{x\}) = 0$ si $x \notin D$.

Así que, si $a < b$:

$\mu_{F^d}((a, b]) = F^d(b) - F^d(a) = \sum_{\{y \in D: a < y \leq b\}} \left[F^d(y) - F^d(y-)\right] = \sum_{\{y \in D: a < y \leq b\}} \mu_{F^d}(\{y\}) = \sum_{y \in (a, b]} \mu_{F^d}(\{y\})$

Por otra parte, si D es finito, sea B un conjunto infinito numerable contenido en $\mathbb{R} - D$, definamos $E = D \cup B$ y denotemos por $y_1, y_2, \ldots$ a los elementos de E. Obviamente se tiene:

$\mu_{F^d}((a, b]) = \sum_{\{y \in E: a < y \leq b\}} \mu_{F^d}(\{y\})$

Para cada subconjunto A de $\mathbb{R}$, definamos:

$\mu(A) = \sum_{\{y \in A \cap E\}} \mu_{F^d}(\{y\})$

Por el Corolario 3.9, μ es una medida definida sobre la σ-álgebra formada por todos los subconjuntos de $\mathbb{R}$.

Además, por el teorema de Clases Monótonas, μ y μ_{F^d} coinciden sobre la σ-álgebra de los conjuntos borelianos de $\mathbb{R}$.

Resumiento, si $F : \mathbb{R} \mapsto \mathbb{R}$ es una función no decreciente, no constante, continua por la derecha y que crece únicamente mediante saltos, entonces existe un conjunto finito o infinito numerable E tal que $\mu_F(\{y\}) > 0$ para cualquier $y \in E$ y, si B es un subconjunto boreliano de $\mathbb{R}$, entonces:

$\mu_F(B) = \sum_{\{y \in B \cap E\}} \mu_F(\{y\})$

En este caso, μ_F puede extenderse a una medida definida sobre la σ-álgebra formada por todos los subconjuntos de $\mathbb{R}$.

Como podemos ver, la medida μ_F se concentra en los elementos de E; en otras palabras, $\mu_F(E^c) = 0$.

Una medida μ de este tipo, para la cual existe un conjunto de números reales, finito o infinito numerable, E tal que $\mu_F(\{y\}) > 0$ para cualquier $y \in E$ y $\mu(E^c) = 0$ es llamada medida discreta. Esto no implica que los elementos de E estén separados topológicamente en $\mathbb{R}$; es decir, no necesariamente ocurre que para cada $y \in E$ existe un intervalo abierto que contiene a y y que no contiene a ningún otro elemento de E. Incluso el conjunto E puede ser denso en $\mathbb{R}$; por ejemplo, sea $\mathbb{Q} = \{r_1, r_2, \ldots\}$ el conjunto formado por todos los números racionales y, para cada subconjunto A de $\mathbb{R}$, definamos:

$\mu(A) = \sum_{\{n \in \mathbb{N}: r_n \in A\}} \frac{1}{2^n}$

Entonces μ es una medida discreta concentrada en $\mathbb{Q}$. Una función no decreciente y continua por la derecha $F : \mathbb{R} \mapsto \mathbb{R}$ que genera tal medida está dada por:

$F(x) = \sum_{\{n \in \mathbb{N}: r_n \leq x\}} \frac{1}{2^n}$

Esta función F crece únicamente mediante saltos y $\mathbb{Q}$ es el conjunto de puntos donde es discontinua.

6.3. Medidas y funciones no decrecientes continuas

Consideremos ahora una función $F : \mathbb{R} \mapsto \mathbb{R}$, no decreciente, no constante y que no tiene discontinuidades.

En este caso, la medida μ_F de cualquier intervalo acotado, con extremos a y b, es igual a $F(b) - F(a)$ y $\mu_F(\{x\}) = 0$ para cualquier $x \in \mathbb{R}$. Además, como F es no constante, existen $a, b \in \mathbb{R}$ tales que $a < b$ y $F(b) - F(a) > 0$.

Por ser F continua, este caso parece más simple que el anterior, donde F es una función que crece únicamente mediante saltos. Sin embargo, en realidad el caso en que F es continua, en general, es más complicado. Esto se puede analizar desde diferentes puntos de vista; por ejemplo, las funciones continuas no siempre son tan "bonitas" como uno las imagina. Las que uno dibuja habitualmente son funciones que no únicamente son continuas, sino que además son derivables en casi todo punto. Pero sabemos que hay funciones continuas $f : [0, 1] \mapsto \mathbb{R}$ que no son derivables en ningún punto $x \in [0, 1]$; de hecho, en un sentido, existen muchas más funciones de este tipo que las que son derivables en algún punto (cf. Bachman G. & Narici L., Functional Analysis, Sección 6.3, Ed. Academic Press, 1966).

El problema de determinar las propiedades que debe tener una función F, no decreciente y continua, para que la medida μ_F que genera sea más fácil de trabajar lo trató Lebesgue en su libro de 1904. Ahí no estudió únicamente el problema de la integración de funciones, sino también el de la búsqueda de primitivas de una función (i.e. dada una función f, encontrar una función cuya derivada sea f), el cual abordó introduciendo el concepto de integral indefinida: Si $f : [a,b] \to \mathbb{R}$ es una función medible e integrable, la función $F : [a,b] \to \mathbb{R}$ definida mediante la relación $F(x) = \int_a^x f(y)\,dy + K$, donde K es una constante, es llamada una integral indefinida de f. Demostró entonces que si F es una integral indefinida de f, entonces F es continua, de variación acotada y $F'(x) = f(x)$ excepto a lo más en un conjunto de medida cero. Probó además que existen funciones continuas y no decrecientes que no son integrales indefinidas de ninguna otra función (ver la introducción al capítulo 5). Poco tiempo después, Vitali y el mismo Lebesgue demostraron que el concepto de integral indefinida coincide con el de función absolutamente continua: Una función $F : [a,b] \to \mathbb{R}$ es absolutamente continua si, dada $\varepsilon > 0$, existe $\delta > 0$ tal que $\sum_{k=1}^{n} |F(b_k) - F(a_k)| < \varepsilon$ para cualquier colección finita de intervalos (a_1, b_1), (a_2, b_2), ..., (a_n, b_n), contenidos en $[a,b]$, ajenos por parejas y tales que $\sum_{k=1}^{n} (b_k - a_k) < \delta$ (cf. Royden, H.L., Real Analysis, second edition, cap. 5, sección 4, Ed. Macmillan, 1968). Estas ideas y resultados condujeron a uno de los teoremas más importantes de la teoría de integración de Lebesgue, el de Radon-Nikodym (ver capítulo 6).

Resumiento, si $F : \mathbb{R} \mapsto \mathbb{R}$ es una función no decreciente, no constante y continua, diremos que F es absolutamente continua, con respecto a la medida de Lebesgue, si existe una función $f : \mathbb{R} \mapsto \mathbb{R}$, no negativa e integrable tal que:

$F(x) = \int_{-\infty}^{x} f(y)\,dy$

La medida μ_F que genera una tal función está dada por:

$\mu_F(B) = \int_B f(y)\,dy$

para cualquier subconjunto boreliano B de $\mathbb{R}$.

Una medida de este tipo no puede extenderse a una medida definida sobre todos los subconjuntos de $\mathbb{R}$.

Una función no decreciente, no constante y continua $F : \mathbb{R} \mapsto \mathbb{R}$ que no sea absolutamente continua, es llamada singular. Las medidas generadas por las funciones singulares son las más dificiles de tratar.

6.4. Medidas con signo y funciones de variación acotada

TEOREMA 6.6. *Si $g : \mathbb{R} \mapsto \mathbb{R}$ es una función de variación acotada sobre cualquier intervalo compacto la cual se puede expresar como la diferencia de dos funciones no decrecientes, una de las cuales, por lo menos, genera una medida finita sobre los conjuntos borelianos de $\mathbb{R}$, entonces existe una única medida con signo ν_g sobre los conjuntos borelianos de $\mathbb{R}$ tal que:*

(i) $\nu_g((a,b]) = g(b+) - g(a+)$,

(ii) $\nu_g([a,b)) = g(b-) - g(a-)$,
(iii) $\nu_g(\{x\}) = g(x+) - g(x-)$,

para cualquier $x \in \mathbb{R}$ y cualquier pareja de números reales, a y b, tales que $a < b$.

Demostración

Sea $g : \mathbb{R} \to \mathbb{R}$ una función de variación acotada que satisface las condiciones del enunciado y sean $f_1 : \mathbb{R} \to \mathbb{R}$ y $f_2 : \mathbb{R} \to \mathbb{R}$ dos funciones no decrecientes tales que $g = f_1 - f_2$ y, por lo menos una de ellas, genera una medida finita sobre los conjuntos borelianos de $\mathbb{R}$.

Sean μ_{f_1} y μ_{f_2} la medidas sobre los conjuntos borelianos de $\mathbb{R}$ generadas por f_1 y f_2, respectivamente. Entonces:

$\mu_{f_1}((a,b]) = f_1(b+) - f_1(a+)$,

$\mu_{f_1}([a,b)) = f_1(b-) - f_1(a-)$,

$\mu_{f_2}((a,b]) = f_2(b+) - f_2(a+)$,

$\mu_{f_2}([a,b)) = f_2(b-) - f_2(a-)$,

para cualquier pareja de números reales, a y b, tales que $a < b$.

Además, $\mu_{f_1}(\{x\}) = f_1(x+) - f_1(x-)$ y $\mu_{f_2}(\{x\}) = f_2(x+) - f_2(x-)$ para cualquier $x \in \mathbb{R}$.

Definamos para cada $E \in \mathfrak{B}(\mathbb{R})$:

$\mu_g(E) = \mu_{f_1}(E) - \mu_{f_2}(E)$.

Si $x \in \mathbb{R}$ y a y b son dos números reales tales que $a < b$, se tiene:

$\mu_g((a,b]) = g(b+) - g(a+)$,

$\mu_g([a,b)) = g(b-) - g(a-)$,

$\mu_g(\{x\}) = g(x+) - g(x-)$.

Además μ_g es una medida con signo ya que si $A_1, A_2, \ldots$son elementos de $\mathfrak{B}(\mathbb{R})$ tales que $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces:

$$\mu_g\left(\bigcup_{k=1}^{\infty} A_k\right) = \mu_{f_1}\left(\bigcup_{k=1}^{\infty} A_k\right) - \mu_{f_2}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty}\left(\mu_{f_1}(A_k) - \mu_{f_2}(A_k)\right)$$

$$= \sum_{k=1}^{\infty}\left(\mu_{f_1} - \mu_{f_2}\right)(A_k) = \sum_{k=1}^{\infty}\mu_g(A_k).$$

Si $\mu_g\left(\bigcup\limits_{k=1}^{\infty} A_k\right) \in \mathbb{R}$, entonces, como por lo menos una de las dos medidas μ_{g_1} y μ_{g_2} es finita, $\mu_{g_1}\left(\bigcup\limits_{k=1}^{\infty} A_k\right)$ y $\mu_{g_2}\left(\bigcup\limits_{k=1}^{\infty} A_k\right)$ son finitas, así que las series $\sum_{k=1}^{\infty}\mu_{g_1}(A_k)$ y $\sum_{k=1}^{\infty}\mu_{g_2}(A_k)$ son convergentes. Finalmente, se tiene:

$\sum_{k=1}^{\infty}\left|\mu_g(A_k)\right| = \sum_{k=1}^{\infty}\left|\mu_{g_1}(A_k) - \mu_{g_2}(A_k\right| \leq \sum_{k=1}^{\infty}\left|\mu_{g_1}(A_k)\right| + \sum_{k=1}^{\infty}\left|\mu_{g_2}(A_k)\right|$.

Así que la serie $\sum_{k=1}^{\infty}\mu(A_k)$ es absolutamente convergente.

Sean ν_1 y ν_2 son dos medidas con signo sobre los borelianos de $\mathbb{R}$ tales que:

$\nu_1\left((a,b]\right) = g\left(b+\right) - g\left(a+\right)$,

$\nu_2\left((a,b]\right) = g\left(b+\right) - g\left(a+\right)$,

para cualquier pareja de números reales, a y b, tales que $a < b$.

Definamos, para cada $n \in \mathbb{N}$, $F_n = (-n, n]$; entonces $\bigcup_{n=1}^{\infty} F_n = \mathbb{R}$, y, para cualquier $n \in \mathbb{N}$, $\nu_1(F_n) = \nu_2(F_n) < \infty$ y $\nu_1(F_n \cap (a,b]) = \nu_2(F_n \cap (a,b])$ cualquier pareja de números reales, a y b, tales que $a < b$. Así que, por el teorema de clases monótonas para π-sistemas, $\nu_1(A) = \nu_2(A)$ para cualquier conjunto boreliano de $\mathbb{R}$.

■

DEFINICIÓN 6.2. *Si $g : \mathbb{R} \mapsto \mathbb{R}$ es una función de variación acotada sobre cualquier intervalo compacto la cual se puede expresar como la diferencia de dos funciones no decrecientes, una de las cuales, por lo menos, genera una medida finita sobre los conjuntos borelianos de $\mathbb{R}$, la medida ν_g del teorema anterior será llamada la medida generada por g.*

TEOREMA 6.7. *Para cualquier medida con signo ν sobre los conjuntos borelianos de $\mathbb{R}$, tal que los intervalos acotados tienen medida finita, existe una función de variación acotada sobre cualquier intervalo compacto $g_\nu : \mathbb{R} \mapsto \mathbb{R}$ la cual se puede expresar como la diferencia de dos funciones no decrecientes, una de las cuales, por lo menos, genera una medida finita sobre los conjuntos borelianos de $\mathbb{R}$, y tal que:*

(i) $\nu\left((a,b]\right) = g_\nu\left(b+\right) - g_\nu\left(a+\right)$,
(ii) $\nu\left([a,b)\right) = g_\nu\left(b-\right) - g_\nu\left(a-\right)$,
(iii) $\nu\left(\{x\}\right) = g_\nu\left(x+\right) - g_\nu\left(x-\right)$,
para cualquier $x \in \mathbb{R}$ y cualquier pareja de números reales, a y b, tales que $a < b$.

Además, si $g_1 : \mathbb{R} \mapsto \mathbb{R}$ y $g_2 : \mathbb{R} \mapsto \mathbb{R}$ son dos funciones con estas propiedades, entonces las funciones $x \to (g_1 - g_2)(x+)$ y $x \to (g_1 - g_2)(x-)$ son constantes.

Demostración

Sea ν una medida con signo definida sobre los conjuntos borelianos de $\mathbb{R}$. Existen entonces dos medidas ν^+ y ν^- sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ con las siguientes propiedades:

1. Por lo menos una de las dos medidas, ν^+ y ν^-, es finita.

2. $\nu = \nu^+ - \nu^-$.

3. Existen dos conjuntos $A, B \in \Im$ tales que $A \cap B = \emptyset$, $\mathbb{F} = A \cup B$ y $\nu^-(A) = \nu^+(B) = 0$.

Supongamos que los intervalos acotados tienen medida ν finita, entonces ν^+ y ν^- tienen la misma propiedad. Así que existen dos funciones no decrecientes $F_{\nu^+} : \mathbb{R} \mapsto \mathbb{R}$ y $F_{\nu^-} : \mathbb{R} \mapsto \mathbb{R}$, tales que:

$\nu^+((a,b]) = F_{\nu^+}(b+) - F_{\nu^+}(a+)$,

$\nu^+([a,b)) = F_{\nu^+}(b-) - F_{\nu^+}(a-)$,

$\nu^-((a,b]) = F_{\nu^-}(b+) - F_{\nu^-}(a+)$,

$\nu^-([a,b)) = F_{\nu^-}(b-) - F_{\nu^-}(a-)$,

para cualquier pareja de números reales, a y b, tales que $a < b$.

Si definimos $g_\nu = F_{\nu^+} - F_{\nu^-}$, entonces g_ν es de variación acotada sobre cualquier intervalo compacto y se tiene:

$\nu((a,b]) = \nu^+((a,b]) - \nu^-((a,b]) = (F_{\nu^+} - F_{\nu^-})(b+) - (F_{\nu^+}(a-) - F_{\nu^-})(a+)$

$= g_\nu(b+) - g_\nu(a+)$,

$\nu([a,b)) = \nu^+([a,b)) - \nu^-([a,b)) = (F_{\nu^+} - F_{\nu^-})(b-) - (F_{\nu^+}(a-) - F_{\nu^-})(a-)$

$= g_\nu(b-) - g_\nu(a-)$,

para cualquier pareja de números reales, a y b, tales que $a < b$.

Además, como la función $g_\nu^d : \mathbb{R} \to \mathbb{R}$ definida por $g_\nu^d = g_\nu(x+)$ es continua por la derecha, y la función $g_\nu^d : \mathbb{R} \to \mathbb{R}$ definida por $g_\nu^i = g_\nu(x-)$ es continua por la izquierda, se tiene:

$\mu(\{x\}) = \text{lím}_{n \rightsquigarrow \infty} \mu\left(\left(x - \frac{1}{n}, x\right]\right) = \text{lím}_{n \rightsquigarrow \infty} \left[g_\nu(x+) - g_\nu^d\left(x - \frac{1}{n}\right)\right]$

$= g_\nu(x+) - g_\nu^d(x-) = g_\nu(x+) - g_\nu(x-)$,

para cualquier $x \in \mathbb{R}$.

Si g_1 y g_2 son dos funciones de variación acotada sobre cualquier intervalo compacto con estas propiedades, entonces $g_1(b+) - g_1(a+) = g_2(b+) - g_2(a+)$ para cualquier pareja de números reales a y b tales que $a < b$, lo cual implica que $(g_1 - g_2)(b+) = (g_1 - g_2)(a+)$. Así que $(g_1 - g_2)(x+)$ es constante.

También se tiene $g_1(b-) - g_1(a-) = g_2(b-) - g_2(a-)$ para cualquier pareja de números reales a y b tales que $a < b$, lo cual implica que $(g_1 - g_2)(b-) = (g_1 - g_2)(a-)$. Así que $(g_1 - g_2)(x-)$ es constante.

■

CAPÍTULO 7

TEORÍA GENERAL DE INTEGRACIÓN
Primera parte

7.1. Introducción

Como lo mencionamos con anterioridad, en el surgimiento y desarrollo de la Teoría de la Medida, el problema central que se abordó no fue el de la medida de conjuntos en sí mismo, sino el de la integración de funciones. Tanto Lebesgue como quienes le siguieron buscaron resolver problemas que tenían que ver con la definición y las propiedades de la integral.

Después que Lebesgue formuló su definición de integral, él mismo y otros autores fueron aportando más ideas y resultados, los cuales iban conformando el cuerpo de una teoría para la cual también se iban encontrando aplicaciones.

Uno de los grandes promotores de la teoría de integración de Lebesgue fue Pierre Joseph Louis Fatou, quien en el año 1906 obtuvo su doctorado con una tesis titulada *Séries trigonométriques et séries de Taylor* ([**32**]), donde utilizó la teoría de Lebesgue para el estudio de la integral de Poisson de una función discontinua en la frontera de la región donde está definida y para tratar problemas relativos al desarrollo de una función en serie trigonométrica. En este trabajo demostró el resultado conocido ahora como Lema de Fatou (el cual se demostrará más adelante). Además de aportar los resultados originales que se encuentran en su tesis, Fatou contribuyó de manera importante al desarrollo de la teoría de integración por la influencia que tuvo su trabajo en el mismo Lebesgue y sobre todo en F. Riesz, quien en un artículo de 1949, titulado *Lévolution de la notion d'integrale depuis Lebesgue* ([**82**]) dijo:

> *"Si no me equivoco, es el libro de Lebesgue sobre las series trigonométricas, dentro de la Colección Borel, el que llamó mi atención sobre la noción de integral; después, para penetrar en los detalles, estudié también su Tesis y su libro sobre la integración. Sin embargo, la idea y el coraje para tratar de aplicar esta noción a los problemas de los que yo me ocupaba, me vinieron leyendo, en 1906, la excelente Memoria de Fatou, impresa en las Acta Mathematica y que el autor presentaba también como Tesis. Fue en particular un teorema muy simple, llamado generalmente lema de Fatou y que asegura, en el lenguaje actual, la semicontiuidad inferior de la operación funcional lineal que constituye la integración, el que me ayudó a demostrar, en febrero de 1907, algunas semanas después de la lectura de la Tesis, el teorema descubierto también, de manera independiente y simultáneamente, por Ernest Fischer y que se cita con el nombre de nosotros dos. El teorema sirvió, en primer lugar, de boleto permanente de ida y vuelta entre los dos espacios con una infinidad de dimensiones cuyo interés se liga con la teoría de las ecuaciones integrales, a saber, el espacio con una infinidad de coordenadas de Hilbert y el conjunto L^2 de las funciones medibles y de cuadrado integrable, dos espacios que, por cierto, actualmente se tratan, con Von Neumann, como dos realizaciones de una noción más general, a saber, el espacio abstracto de Hilbert. Fue quizás la primera aplicación de la teoría de Lebesgue, después, bien entendido, de las que fueron hechas por él mismo y por Fatou, la que atrajo el interés de los matemáticos y que daba luz sobre la importancia de su noción de integral."*

Otro de los grandes promotores de la nueva teoría de integración fue precisamente el autor de la frase del párrafo anterior, Frédéric Riesz, quien aplicó la nueva teoría de integración de Lebesgue al Análisis Funcional.

Uno de los resultados más importantes de Lebesgue es el que se refiere a la segunda parte del título de su libro. Recordemos que el libro de Lebesgue de 1904 tiene como título *Leçons sur l'intégration et la recherche des fonctions primitives.* Hasta el momento, en este texto, hemos hablado de la medida e integral de Lebesgue y hemos dejado de lado el tema de la búsqueda de funciones primitivas, es decir, dada una función f, determinar, si existe, una función cuya derivada sea f. Las investigaciones alrededor de este problema culminaron con un artículo de Otto Nikodym, publicado en 1930 con el título *Sur une généralisation des intégrales de M. J. Radon* ([**73**]), en el cual demostró el ahora llamado teorema de Radon-Nikodym, resultado que permitió definir de manera general un concepto de importancia central en la teoría de los procesos estocásticos, el de Esperanza Condicional.

El tema de la búsqueda de funciones primitivas lo abordó Lebesgue con el estudio de las integrales indefinidas:

Si $f : [a, b] \to \mathbb{R}$ es una función medible e integrable, la función $F : [a, b] \to \mathbb{R}$ definida mediante la relación $F(x) = \int_a^x f(y)\,dy + K$, donde K es una constante, era llamada por Lebesgue una integral indefinida de f.

Lebesgue demostró que las integrales indefinidas tienen las siguientes tres propiedades:

(i) Son funciones continuas.
(ii) Son de variación acotada.
(iii) Tienen como derivada a la función de la cual es una integral indefinida, excepto a lo más en los puntos de un conjunto de medida cero.

El estudio que hizo Lebesgue sobre este tema en su libro fue incompleto, se dieron más tarde resultados de otros autores que fueron completando el cuadro. Sin embargo, al final del libro Lebesgue introdujo una propiedad que sería clave para tratar el problema de la relación entre la integral y la derivada; propiedad que, además, llevaría a uno de los resultados más importantes de su teoría de integración, el cual, a su vez, sería una de las bases para el estudio de los procesos estocásticos, tema que trataremos más adelante.

Después de una serie de razonamientos, concluía Lebesgue:

> *"Queda así demostrado que toda función de variación acotada $f(x)$ tiene una derivada finita excepto para los valores de x de un conjunto de medida cero [resultado importante en sí mismo]. El razonamiento de la página* 122*, tal como acaba de ser completado, muestra también que esta derivada es integrable en el conjunto de puntos donde es finita, pero su función primitiva no necesariamente es $f(x)$, como lo muestra el ejemplo de la función $\xi(x)$ de la página* 55*. El teorema que acaba de ser demostrado es por consiguiente diferente del que concierne a la derivación de las integrales indefinidas; en otros términos, existen funciones continuas de variación acotada, $\xi(x)$ por ejemplo, que no son integrales indefinidas."*

El ejemplo al que se refería Lebesgue es el siguiente:

Sea C el conjunto de Cantor, entonces cada $x \in C$ se puede expresar como una serie:

$x = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \cdots,$

donde $a_k \in \{0, 2\}$ para cualquier $k \in \mathbb{N}$.

Para cada $x = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \cdots \in C$, definamos:

$\xi(x) = \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \cdots\right).$

ξ es no decreciente ya que si $x, y \in C$ y $x < y$, entonces, si, en los desarrollos en base 3 de x y y, el m-simo es el primero que es distinto, ese término tiene que ser 0 para x y 2 para y; así que, si $a_1, a_2, \ldots, a_{m-1}$ son los $m-1$términos de los desarrollos de x y y, se tiene:

$\xi(x) = \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \cdots + \frac{a_{m-1}}{2^{m-1}}\right) + \frac{1}{2}\sum_{k=m+1}^{\infty} \frac{a_k}{2^k}$,

$\xi(y) = \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \cdots + \frac{a_{m-1}}{2^{m-1}}\right) + \frac{1}{2^m} + \frac{1}{2}\sum_{k=m+1}^{\infty} \frac{b_k}{2^k}$.

Por lo tanto:

$\xi(x) \leq \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \cdots + \frac{a_{m-1}}{2^{m-1}}\right) + \frac{1}{2}\sum_{k=m+1}^{\infty} \frac{2}{2^k}$

$= \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \cdots + \frac{a_{m-1}}{2^{m-1}}\right) + \sum_{k=m+1}^{\infty} \frac{1}{2^k} = \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \cdots + \frac{a_{m-1}}{2^{m-1}}\right) + \frac{1}{2^m}$,

$\xi(y) \geq \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \cdots + \frac{a_{m-1}}{2^{m-1}}\right) + \frac{1}{2^m} \geq \xi(x)$.

Así que, las discontinuidades de ξ únicamente pueden ser de saltos; pero al ser ξ suprayectiva como función de C en el intervalo $[0,1]$, no puede tener saltos. Por lo tanto, ξ es continua.

Para definir ξ en todo el intervalo $[0,1]$, falta definirla en los intervalos abiertos que se suprimen del intervalo $[0,1]$ para formar C.

Los desarrollos en base 3 de los extremos de un intervalo que se suprime en el n-simo paso coinciden hasta el término $n-1$, y el término siguiente del extremo izquierdo del intervalo es cero, mientras que el del extremo derecho es 2. Así que, si (c, d) es uno de los intervalos que se suprimen en el paso n, se tiene:

$c = 0.a_1 a_2 \cdots a_{n-1} 0222 \cdots$

$d = 0.a_1 a_2 \cdots a_{n-1} 2000 \cdots$

Así que:

$\xi(c) = \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_{n-1}}{2^{n-1}}\right) + \frac{1}{2}\sum_{k=n+1}^{\infty} \frac{2}{2^k} = \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_{n-1}}{2^{n-1}}\right) + \frac{\frac{1}{2^{n+1}}}{\frac{1}{2}}$

$= \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_{n-1}}{2^{n-1}}\right) + \frac{1}{2^n}$,

$\xi(d) = \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_{n-1}}{2^{n-1}} + \frac{2}{2^n}\right) = \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_{n-1}}{2^{n-2}}\right) + \frac{1}{2^n}$.

Por lo tanto, $\xi(c) = \xi(d)$.

Definamos $\xi(x) = \xi(c)$ para cualquier $x \in (c, d)$.

ξ es entonces una función continua y no decreciente, definida sobre el intervalo $[0,1]$ y con valores en el mismo intervalo.

Sean $(c_1, d_1), (c_2, d_2), \ldots$ los intervalos que se suprimen para formar el conjunto de Cantor, entonces:

$[0,1] = C \cup (\bigcup_{k=1}^{\infty} (c_k, d_k))$.

Supongamos que ξ es una integral indefinida de una función medible e integrable $f : [0,1] \to \mathbb{R}$, entonces existe una constante $K \in \mathbb{R}$ tal que:

$\xi(x) = \int_0^x f(y)\, dy + K$,

para cualquier $x \in [0,1]$.

ξ es derivable y su derivada es cero en cualquier punto del conjunto $D = \bigcup_{k=1}^{\infty} (c_k, d_k)$, el cual tiene medida de Lebesgue 1.

Por otra parte, al ser ξ una integral indefinida de f, su derivada es f, excepto a lo más en los puntos de un conjunto de medida cero.

Sean:

$A = \{x \in (0,1) : \xi'(x) \text{ existe y } \xi'(x) \neq f(x)\}$,

$B = \{x \in (0,1) : \xi'(x) \text{ no existe o } (\xi'(x) \text{ existe y } \xi'(x) = f(x))\}$.

A tiene entonces medida de Lebesgue cero y B, que es el complemento de A en el intervalo $(0,1)$, tiene medida de Lebesgue 1.

Por lo tanto, $B \cap D$ tiene medida de Lebesgue 1. Además:

$B \cap D = \{x \in D : \xi'(x) = f(x)\}$.

Así que $f(x) = 0$ para cualquier $x \in B \cap D$.

Por lo tanto $f = 0$ excepto a lo más en un conjunto de medida de Lebesgue cero.

Se tiene entonces:

$\int_0^x f(y)\, dy = 0$ para cualquier $x \in [0,1]$.

ξ tendría entonces que ser constante, pero no lo es, lo cual es una contradicción.

Por lo tanto, ξ no es una integral indefinida.

Al final de su libro, Lebesgue agregó una nota después de que afirma que existen funciones continuas de variación acotada que no son integrales indefinidas. Lo que afirma, sin demostración, en esa nota es de una importancia central para el desarrollo de un tema que conduciría al teorema de Radon-Nykodim, del cual hablaremos más adelante y lo expondremos formalmente en este capítulo.

La nota de Lebesgue dice:

> *"Para que una función sea integral indefinida, se requiere además que su variación total en una infinidad numerable de intervalos de longitud total ℓ, tienda hacia cero con ℓ."*

En otras palabras, para que $F : [a,b] \to \mathbb{R}$ sea una integral indefinida se requiere que, si $(a_1, b_1), (a_2, b_2), (a_3, b_3), \ldots$ son intervalos contenidos en $[a,b]$, ajenos por parejas, entonces:

$\lim_{\sum_{k=1}^{\infty}(b_k - a_k)\to 0} \sum_{k=1}^{\infty} |F(b_k) - F(a_k)| = 0.$

Es decir, dada $\varepsilon > 0$, existe $\delta > 0$ tal que $\sum_{k=1}^{\infty} |F(b_k) - F(a_k)| < \varepsilon$ para cualquier sucesión de intervalos ajenos por parejas $((a_k, b_k))_{k\in\mathbb{N}}$ contenidos en $[a,b]$ y tales que $\sum_{k=1}^{\infty}(b_k - a_k) < \delta$.

Esta propiedad es equivalente a la siguiente:

Dada $\varepsilon > 0$, existe $\delta > 0$ tal que $\sum_{k=1}^{n} |F(b_k) - F(a_k)| < \varepsilon$ para cualquier colección finita de intervalos (a_1, b_1), (a_2, b_2), $\ldots$, (a_n, b_n), contenidos en $[a,b]$, ajenos por parejas y tales que $\sum_{k=1}^{n}(b_k - a_k) < \delta$.

En efecto, supongamos que se tiene la primera propiedad y, dada $\varepsilon > 0$, tomemos $\delta \in (0, b-a)$ tal que $\sum_{k=1}^{\infty} |F(b_k) - F(a_k)| < \varepsilon$ para cualquier sucesión de intervalos ajenos $([a_k, b_k])_{k\in\mathbb{N}}$ contenidos en $[a,b]$ y tales que $\sum_{k=1}^{\infty}(b_k - a_k) < \delta$. Entonces, dada cualquier colección finita de intervalos (a_1, b_1), (a_2, b_2), $\ldots$, (a_n, b_n), contenidos en $[a,b]$, ajenos por parejas y tales que $\sum_{k=1}^{n}(b_k - a_k) < \delta$, podemos agregar a esa familia una colección infinita numerables de intervalos $(a_{n+1}, bn{+}_1)$, (a_{n+2}, b_{n+2}), $\ldots$, contenidos en $[a,b]$, ajenos por parejas, sin puntos en común con los primeros n intervalos y tales que $\sum_{k=n+1}^{\infty}(b_k - a_k) < \delta - \sum_{k=1}^{n}(b_k - a_k)$; así que $\sum_{k=1}^{\infty}(b_k - a_k) < \delta$ y entonces:

$\sum_{k=1}^{n} |F(b_k) - F(a_k)| \leq \sum_{k=1}^{\infty} |F(b_k) - F(a_k)| < \varepsilon.$

Inversamente, supongamos que se tiene la segunda propiedad y, dada $\varepsilon > 0$, tomemos $\delta > 0$ tal que $\sum_{k=1}^{n} |F(b_k) - F(a_k)| < \frac{1}{2}\varepsilon$ para cualquier colección finita de intervalos (a_1, b_1), (a_2, b_2), $\ldots$, (a_n, b_n), contenidos en $[a,b]$, ajenos por parejas y tales que $\sum_{k=1}^{n}(b_k - a_k) < \delta$. Consideremos una sucesión de intervalos ajenos por parejas $((a_k, b_k))_{k\in\mathbb{N}}$ contenidos en $[a,b]$ y tales que $\sum_{k=1}^{\infty}(b_k - a_k) < \delta$; entonces, para cualquier $n \in \mathbb{N}$, los intervalos (a_1, b_1), (a_2, b_2), $\ldots$, (a_n, b_n) están contenidos en $[a,b]$, son ajenos por parejas y $\sum_{k=1}^{n}(b_k - a_k) < \delta$; por lo tanto, $\sum_{k=1}^{n} |F(b_k) - F(a_k)| < \frac{1}{2}\varepsilon$. Así que:

$\sum_{k=1}^{\infty} |F(b_k) - F(a_k)| = \lim_{n\rightsquigarrow\infty} \sum_{k=1}^{n} |F(b_k) - F(a_k)| \leq \frac{1}{2}\varepsilon < \varepsilon.$

La segunda propiedad es la definición moderna de una función absolutamente continua. Así que, lo que afirmó Lebesgue es que, para que una función de variación acotada sea una integral indefinida se requiere que sea absolutamente continua. Agregó, en la misma nota, también sin demostración, que esta condición no únicamente es necesaria para que una función sea una integral indefinida, sino que también es suficiente.

En 1905, Giuseppe Vitali publicó una demostración de la afirmación de Lebesgue en un artículo titulado *Sulle funzioni integrali* ([**90**]), extendiendo el resultado al caso multidimensional. Fue Vitali quien dio el nombre de continuidad absoluta a la propiedad enunciada por Lebesgue. Más tarde, en 1907, Lebesgue publicó su propia demostración en un artículo titulado *Sur la recherche des fonctions primitives par l'intégration* ([**59**]).

> En resumen, los resultados de Lebesgue y Vitali, para el caso unidimensional, son los siguientes:
> (i) Una función es una integral indefinida si y sólo si es absolutamente continua.
> (ii) Toda integral indefinida es de variación acotada.
> (iii) Si $F : [a,b] \to \mathbb{R}$ es una integral indefinida de la función f, entonces existe un conjunto A de medida cero tal que $F'(x)$ existe para cualquier $x \notin A$ y $F'(x) = f(x)$.
> (iv) No toda función de variación acotada continua es una integral indefinida. Existen funciones de variación acotadas continuas, no constantes, cuya derivada es cero excepto a lo más en un conjunto de medida de Lebesgue cero, así que tales funciones no son integrales indefinidas.

En 1910 Lebesgue publicó un artículo, titulado *L'intégration des fonctions discontinues* ([**60**]), donde profundizó el estudio de las integrales indefinidas. En ese artículo analizó las integrales definidas para el caso multidimensional, planteando un nuevo enfoque (al parecer, influenciado por un trabajo de Vitali, de 1907-1908, sobre el mismo tema), el cual sería retomado por Johann Radon, en el año 1913, en un artículo que, como ya lo mencionamos en la introducción del capítulo 5, sentó las bases para desarrollar una teoría general de la medida.

En ese mismo artículo de 1910, Lebesgue demostró el ahora conocido como el teorema de la convergencia dominada, el cual afirma que si $(f_n)_{n\in\mathbb{N}}$ es una sucesión de funciones medibles tales $\lim_{n\rightsquigarrow\infty} f_n$ existe excepto a lo más en un conjunto de medida cero, y $|f_n| \leq g$, donde g es una función medible cuya integral es finita, entonces:

$$\int \lim_{n\rightsquigarrow\infty} f_n(y)\, dy = \lim_{n\to\infty} \int f_n(y)\, dy.$$

El cambio de enfoque de Lebesgue para tratar el tema de las integrales definidas consistió en considerarlas como funciones definidas sobre los conjuntos medibles. Especificamente, consideró una integral indefinida como una función F que asigna a cada conjunto medible E la integral $\int_E f\ (P)\, dP$, donde f es una función medible e integrable y P representa un elemento de $\mathbb{R}^n$. Demostró entonces que una función así definida tiene las siguientes dos propiedades:

(i) Si $(E_n)_{n\in\mathbb{N}}$ es una sucesión de conjuntos medibles tales que $\lim_{n\rightsquigarrow\infty} m(E_n) = 0$, donde m es la medida de Lebesgue en $\mathbb{R}^n$, entonces $\lim_{n\rightsquigarrow\infty} F(E_n) = 0$.

(ii) Si $(E_n)_{n\in\mathbb{N}}$ es una sucesión de conjuntos medibles, ajenos por parejas, entonces: $F\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} F(E_n)$.

En el artículo de Lebesgue, una función que satisface la propiedad 2 es llamada aditiva.

En 1913, Johann Radon ([**76**]) dio un nuevo paso importante alrededor de este problema al plantearlo de una manera más general. Radon retomó el concepto de función aditiva definida sobre una familia de subconjuntos de $\mathbb{R}^n$, el cual había sido definido por Lebesgue en su artículo de 1910, pero el problema de las integrales definidas lo planteó como un problema de la relación entre dos funcionales aditivas de acuerdo con la siguiente definición:

> Sean b y f dos funcionales aditivas definidas las familias de subconjuntos de $\mathbb{R}^n$, T_b y T_f, respectivamente. Se dice que f es de base b si b es no negativa y si, para cualquier conjunto $E \in T_b \cap T_f$, la relación $b(E) = 0$ implica que $f(E) = 0$.

> Bajo una condición adicional mostró que si f es de base b, entonces $T_f \subset T_b$ y demostró el siguiente resultado:

> Si la funcional aditiva f es de base b, entonces existe una función Ψ, integrable con respecto a b, tal que $f(E) = \int_E \Psi db$ para cualquier conjunto $E \in T_f$.

> Otton Nikodym, en su artículo de 1930 ([**73**]), retomó el trabajo de Radon y obtuvo un resultado general, ahora conocido como el teorema de Radon-Nikodym.

> Nikodym hacía referencia en su artículo a la formulación general que hizo Fréchet de la teoría de la medida de Lebesgue, pero modificó un poco la terminología. Consideraba una familia no vacía $\mathcal{H}$ de subconjuntos de un conjunto H, la cual es cerrada bajo uniones numerables y complementos (en particular H pertenece a la familia); es decir, lo que ahora se denomina σ-álgebra de subconjuntos de H. Una medida μ la definió entonces como una función (con valores reales), no negativa, definida sobre $\mathcal{H}$, la cual es "perfectamente aditiva", es decir, si $E_1, E_2, \ldots$ son elementos de la familia, ajenos por parejas, entonces $\mu\left(\bigcup_{n=1}^{\infty} En\right) = \sum_{n=1}^{\infty} \mu(En)$; es decir, μ es σ-aditiva, en la terminología moderna. En otras palabras, Nikodym trabajaba con medidas tal y como las hemos definido en el capítulo 4 ($\mu(\emptyset) = 0$ se sigue de la σ-aditividad).

Dada una medida μ sobre $\mathcal{H}$, definió la μ-distancia entre dos elementos E y F de $\mathcal{H}$ de la siguiente manera:
$||E, F||_{\mu} = \mu(E - F) + \mu(F - E)$.
Si ν es una función con valores reales definida sobre $\mathcal{H}$, llamaba a ν μ-continua si para cualquier sucesión $(E_n)_{n\in\mathbb{N}}$ de elementos de $\mathcal{H}$ tales que:
$\text{lím}_{n\rightsquigarrow\infty} ||E_n, E||_{\mu} = 0$,
donde $E \in \mathcal{H}$, se tiene $\text{lím}_{n\rightsquigarrow\infty} \nu(E_n) = \nu(E)$.

Después de que desarrolló la teoría de integración con respecto a una medida μ, demostró el resultado central de su artículo:
Si ν es perfectamente aditiva, entonces las 4 condiciones siguientes son equivalentes:

(i) ν es μ-continua.
(ii) Para cualquier conjunto $E \in \mathcal{H}$, si $\nu(E) \neq 0$ entonces $\mu(E) > 0$.
(iii) Si $(E_n)_{n\in\mathbb{N}}$ es una sucesión de elementos de $\mathcal{H}$ tales que $\text{lím}_{n\rightsquigarrow\infty} \mu(E_n) = 0$, entonces $\text{lím}_{n\rightsquigarrow\infty} \nu(E_n) = 0$.
(iv) Existe una función μ-integrable $f : H \to \mathbb{R}$ tal que $\nu(E) = \int_E f d\mu$, para cualquier conjunto $E \in \mathcal{H}$.

Con este trabajo de Nikodym quedó formulada la teoría de la medida como se le conoce actualmente y quedaron establecidos los resultados básicos de la teoría de integración con respecto a una medida, los cuales expondremos en este capítulo.

Pero antes de esa exposición es necesario mencionar que previamente al trabajo de Nikodym, la solución al problema de la integración de funciones, visto como uno de Análisis Funcional, había sido ya formulada de manera completa por Percy John Daniell en sus 4 artículos publicados entre 1918 y 1920 ([**21**], [**22**], [**23**], [**24**]).

En su primer artículo, *A general form of the integral* ([**21**]), publicado en 1918, decía Daniell: "La idea de una integral ha sido extendida por Radon (1913), Young (1914), Riesz (1914) y otros a la integración con respecto a una función de variación acotada. Estas teorías están basadas sobre las propiedades fundamentales de los conjuntos de puntos en un espacio con un número finito de dimensiones. En este artículo se desarrolla una teoría que es independiente de la naturaleza de sus elementos. Pueden ser puntos en un espacio de una infinidad numerable de dimensiones, o curvas en general, o clases de eventos que conciernen a la teoría. Se sigue que, aunque muchas de las demostraciones que se dan [en este artículo] son meras traducciones a otro lenguaje de métodos ya clásicos (particularmente los debidos a Young), aquí y ahí, donde las demostraciones previas se basan en la teoría de conjuntos de puntos, nuevos métodos han sido desarrollados."Mencionaba también que Fréchet consideró una integral general, pero decía que no trató completamente los teoremas de existencia.

Para definir la integral, asumía que hay una clase inicial T_0 de funciones acotadas con valores reales, definidas sobre un conjunto H, la cual tiene las siguientes propiedades:

1. Si $f \in T_0$ y c es una constante, entonces $cf \in T_0$.

2.Si $f_1, f_2 \in T_0$, entonces $f_1 + f_2$, $\text{máx}\,(f_1, f_2)$ y $\text{mín}\,(f_1, f_2)$ pertenecen a T_0.

Consideraba entonces funciones U definidas sobre T_0 (las denominaba funcionales), las cuales pueden tener algunas de las siguientes propiedades:

(A) $U\,(f_1 + f_2) = U\,(f_1) + U\,(f_2)$.

(C) $U\,(cf) = cU\,(f)$, donde c es una constante.

(L) Si $(f_n)_{n\in\mathbb{N}}$ es una sucesión decreciente de funciones tales que $\text{lím}_{n\rightsquigarrow\infty} f_n\,(p) = 0$ para cualquier p, entonces $\text{lím}_{n\to\infty} U\,(f_n) = 0$.

(M) Existe una funcional M, definida sobre las funciones no negativas de T_0, tal que si $\varphi \leq \Psi$, entonces $M\,(\varphi) \leq M\,(\Psi)$, y $|U\,(f)| \leq M\,(|f|)$.

(P) Si f es no negativa, entonces $U\,(f) \geq 0$.

Denominaba I-integral a una funcional I que satisfaga (A), (C), (L) y (P) y S-integral a una funcional que satisfaga (A), (C), (L) y (M). Mencionaba que una I-integral puede ser llamada una integral positiva y que una S-integral es una integral de Stieltjes generalizada.

Como ejemplo mencionaba que si T_0 es el conjunto de funciones continuas definidas sobre un intervalo (a, b), entonces la integral de Riemann es una I-integral y la integral de Stieltjes es una S-integral.

Definió la clase T_1 como la familia de funciones que son límite de una sucesión no decreciente $(f_n)_{n\in\mathbb{N}}$ de elementos de T_0.

Si $f \in T_1$ y $f = \text{lím}_{n\rightsquigarrow\infty} f_n$, define $I\,(f) = \text{lím}_{n\rightsquigarrow\infty} I\,(f_n)$. Si $I\,(f) < \infty$ se dirá que f es sumable.

Si f es cualquier función con valores reales, definida sobre H, se define:

$$\overline{I}\,(f) = \text{ínf}\,\{I\,(\varphi) \in T_1 : \varphi \in T_1 \text{ y } \varphi \geq f\},$$

$$\underline{I}\,(f) = -\overline{I}\,(-f).$$

$$\text{Si } \overline{I}\,(f) = \underline{I}\,(f) < \infty \text{ se dirá que } f \text{ es sumable y se define } I\,(f) = \overline{I}\,(f).$$

Si S es una S- integral, demostró que las funcionales $I_2 = I_1 - S$ e $I = I_1 + I_2$ son I-integrales y se dice que una función f con valores reales, definida sobre H, es sumable (S) si es I-sumable. En este caso, se define $S\,(f) = I_1\,(f) - I_2\,(f)$.

Demostró entonces los siguientes resultados:

(i) I, definida sobre T_1, es una I-integral.

(ii) Si f es el límite de una sucesión no decreciente $(f_n)_{n\in\mathbb{N}}$ de elementos de T_1, entonces $f \in T_1$ e $I(f) = \text{lím}_{n\rightsquigarrow\infty} I(f_n)$.
(iii) I, definida sobre el conjunto de funciones sumables, es una I-integral.
(iv) Si f es el límite de una sucesión no decreciente $(f_n)_{n\in\mathbb{N}}$ de funciones sumables y:
 $\text{lím}_{n\to\infty} I(f_n) < \infty$,
 entonces f es sumable e $I(f) = \text{lím}_{n\rightsquigarrow\infty} I(f_n)$.
(v) Si f es el límite de una sucesión $(f_n)_{n\in\mathbb{N}}$ de funciones sumables y existe una función sumable φ tal que $|f_n| \leq \varphi$ para cualquier $n \in \mathbb{N}$, entonces f es sumable e $I(f) = \text{lím}_{n\to\infty} I(f_n)$.
(vi) S, definida sobre el conjunto de funciones (S) sumables, es una S-integral.
(vii) Si f es el límite de una sucesión no decreciente $(f_n)_{n\in\mathbb{N}}$ de funciones (S) sumables y $\text{lím}_{n\rightsquigarrow\infty} (I_1 + I_2)(f_n) < \infty$, entonces f es (S) sumable y $S(f) = \text{lím}_{n\rightsquigarrow\infty} S(f_n)$.
(viii) Si f es el límite de una sucesión $(f_n)_{n\in\mathbb{N}}$ de funciones (S) sumables y existe una función (S) sumable φ tal que $|f_n| \leq \varphi$ para cualquier $n \in \mathbb{N}$, entonces f es (S) sumable y $S(f) = \text{lím}_{n\rightsquigarrow\infty} S(f_n)$.

Como puede verse, Daniell realizó un proceso de extensión de una funcional lineal, definida sobre un conjunto de funciones T_0, a una funcional lineal definida sobre un conjunto de funciones que es cerrado bajo el paso a límites y la funcional extendida es tal que, bajo determinadas condiciones, la funcional de un límite de funciones es igual al límite de la sucesión formada por la aplicación de la funcional a cada una de las funciones.

Es lo mismo que había realizado Lebesgue, pero el resultado de Daniell no se restringe a funciones definidas sobre $\mathbb{R}^n$; incluye el caso en que las funciones a integrar estén definidas sobre un espacio de dimensión infinita. El teorema de Carathéodory permitió hacer lo mismo siguiendo el enfoque de Lebesgue, basando la definición de la integral en la existencia de una medida. De hecho, el resultado de Carathéodory permite realizar también un proceso de extensión, pero no de una funcional, sino de una función definida sobre conjuntos, para así llegar a la construcción de una medida. La aplicación del teorema de Carathéodory a la construcción de medidas en espacios de dimension infinita tardó algunos años, básicamente se dio con el resultado de Kolmogorov del año 1933.

En el año 1920, Daniell publicó su segundo artículo sobre el tema de la integral, bajo el título *Integrals in an infinity number of dimensions* ([**22**]). En ese trabajo dio algunos ejemplos de integrales de funciones definidas sobre espacios de dimensión infinita. En el mismo año publicó otros dos artículos, uno titulado *Functions of limited variation in an infinite number of dimensions* ([**23**]) y el otro *Further properties of the general integral* ([**24**]), en los cuales continuó desarrollando su teoría de integración.

El trabajo de Daniell tuvo una gran influencia, siendo su principal aplicación la que hizo Norbert Wiener, entre 1921 y 1923 ([**96**], [**97**], [**98**], [**99**], [**102**]) al construir un modelo matemático para el movimiento browniano, no basándose en la teoría de la medida que se estaba desarrollando, sino utilizando el teorema de extensión de Daniell. Algunos años más tarde, al completarse el desarrollo de la teoría de la medida y de la teoría de integración con respecto a una medida, el método de Daniell fue reemplazado por el de Carathéodory.

Sin embargo, no está dicha la última palabra; recientemente se ha retomado el método de Daniell, en particular en el Cálculo Estocástico.

A continuamos expondremos la formulación moderna de la teoría de integración con respecto a una medida.

7.2. Funciones medibles

En lo que resta de este capítulo, $(\mathbb{F}, \Im)$ será un espacio medible fijo y a los conjuntos $E \in \Im$ los llamaremos conjuntos medibles.

Recordemos que Lebesgue demostró que si $f : [a, b] \to \mathbb{R}$ es una función acotada no negativa, entonces el conjunto:

$E = \{(x, y) \in \mathbb{R}^2 : x \in [a, b] \text{ y } y \in [0, f(x)]\}$,

es medible si y sólo si el conjunto $\{x \in [a, b] : f(x) > \alpha\}$ es medible para cualquier $\alpha \in \mathbb{R}$.

De este resultado surgió la definición que dio de función medible: Una función acotada $f : [a, b] \to \mathbb{R}$ es medible si el conjunto $\{x \in [a, b] : f(x) > \alpha\}$ es medible para cualquier $\alpha \in \mathbb{R}$.

Si definimos $\mathcal{H} = \{B \in \mathfrak{B}(\mathbb{R}) : f^{-1}(B) \in \mathfrak{B}([a, b])\}$, se puede ver fácilmente que $\mathcal{H}$ es una σ-algebra de subconjuntos de $\mathbb{R}$; además, si f es medible (de acuerdo con la definición de Lebesgue) $\mathcal{H}$ contiene a todos los intervalos de la forma (α, ∞). Por lo tanto, contiene contiene a la σ-álgebra generada por esos intervalos, es decir los borelianos. Así que $\mathcal{H} = \mathfrak{B}(\mathbb{R})$. De manera que una función $f : [a, b] \to \mathbb{R}$ es medible si y sólo si $f^{-1}(B) \in \mathfrak{B}([a, b])$ para cualquier $B \in \mathfrak{B}(\mathbb{R})$.

Lo anterior motiva que en un caso general se defina la medibilidad de una función de la siguiente manera:

Definición 7.1. *Sea $(\mathbb{E}, \mathfrak{E})$ un espacio medible. Diremos que una función $f : \mathbb{F} \to \mathbb{E}$ es medible si $f^{-1}(B) \in \Im$ para cualquier $B \in \mathfrak{E}$.*

Ahora mostramos que para que una función $f : \mathbb{F} \to \mathbb{E}$ sea medible basta con que $f^{-1}(B) \in \Im$ para cualquier elemento B de una familia de conjuntos que genere la σ-álgebra $\mathfrak{E}$.

Proposición 7.1. *Sean $(\mathbb{F}, \Im)$ y $(\mathbb{E}, \mathfrak{E})$ dos espacios medibles y $\mathcal{A}$ una familia de subconjuntos de $\mathbb{E}$ tal que $\mathfrak{E} = \sigma(\mathcal{A})$. Entonces una función $f : \mathbb{F} \to \mathbb{E}$ es medible si y sólo si $f^{-1}(B) \in \Im$ para cualquier $B \in \mathcal{A}$.*

Demostración

La familia $\{B \subset \mathbb{E} : f^{-1}(B) \in \Im\}$ es una σ-álgebra de subconjuntos de $\mathbb{E}$ la cual contiene a $\mathcal{A}$, por lo tanto contiene a $\sigma(\mathcal{A})$.

■

PROPOSICIÓN 7.2. *Sean* $(\mathbb{E}, \mathfrak{E})$ *y* $(\mathbb{G}, \mathfrak{G})$ *dos espacios medibles y* $f : \mathbb{F} \to \mathbb{E}$, $g : \mathbb{E} \to \mathbb{G}$ *dos funciones medibles, entonces* $g \circ f$ *es medible.*

Demostración

Sea $B \in \mathfrak{G}$, entonces $g^{-1}(B) \in \mathfrak{E}$, así que:

$(g \circ f)^{-1}(B) = (f^{-1} \circ g^{-1})(B) = f^{-1}(g^{-1}(B)) \in \Im$.

■

7.3. Funciones medibles con valores en $\overline{\mathbb{R}}$

La medibilidad de una función $f : \mathbb{F} \to \mathbb{R}$ (resp. $f : \mathbb{F} \to \overline{\mathbb{R}}$) será entendida considerando sobre $\mathbb{R}$ (resp. $\overline{\mathbb{R}}$) la σ-álgebra de los conjuntos borelianos en $\mathbb{R}$ (resp. $\overline{\mathbb{R}}$).

La proposición 2.1 implica el siguiente resultado:

TEOREMA 7.1. *Una función* $f : \mathbb{F} \to \mathbb{R}$ *es medible si y sólo si se cumple cualquiera de las siguientes propiedades:*

(i) $\{x \in \mathbb{F} : f(x) \leq y]\} \in \Im$ *para cualquier* $y \in \mathbb{R}$.
(ii) $\{x \in \mathbb{F} : f(x) < y\} \in \Im$ *para cualquier* $y \in \mathbb{R}$.
(iii) $\{x \in \mathbb{F} : f(x) \in (a, b]\} \in \Im$ *para cualesquiera* $a, b \in \mathbb{R}$.
(iv) $\{x \in \mathbb{F} : f(x) \in [a, b)\} \in \Im$ *para cualesquiera* $a, b \in \mathbb{R}$.
(v) $\{x \in \mathbb{F} : f(x) \in (a, b)\} \in \Im$ *para cualesquiera* $a, b \in \mathbb{R}$.
(vi) $\{x \in \mathbb{F} : f(x) \in [a, b]\} \in \Im$ *para cualesquiera* $a, b \in \mathbb{R}$.

De la misma manera, los resultados del capítulo 2, acerca de los generadores de la σ-álgebra de Borel en $\overline{\mathbb{R}}$, implican el siguiente resultado:

TEOREMA 7.2. *Una función* $f : \mathbb{F} \to \overline{\mathbb{R}}$ *es medible si y sólo si se cumple cualquiera de las siguientes propiedades:*

(i) $\{x \in \mathbb{F} : f(x) \in (-\infty, y]\} \in \Im$ *para cualquier* $y \in \overline{\mathbb{R}}$.
(ii) $\{x \in \mathbb{F} : f(x) \leq y]\} \in \Im$ *para cualquier* $y \in \mathbb{R}$.
(iii) $\{x \in \mathbb{F} : f(x) < y\} \in \Im$ *para cualquier* $y \in \mathbb{R}$.
(iv) $\{x \in \mathbb{F} : f(x) \leq y]\} \in \Im$ *para cualquier* $y \in \overline{\mathbb{R}}$.
(v) $\{x \in \mathbb{F} : f(x) < y\} \in \Im$ *para cualquier* $y \in \overline{\mathbb{R}}$.

Obsérvese que si una función $f : \mathbb{F} \to \overline{\mathbb{R}}$ es medible, entonces los conjuntos $\{x \in \mathbb{F} : f(x) = \infty\}$ y $\{x \in \mathbb{F} : f(x) = -\infty\}$ son medibles. Así que también el conjunto $\{x \in \mathbb{F} : f(x) \in \mathbb{R}\}$ es medible.

PROPOSICIÓN 7.3. *Toda función* $\varphi : \mathbb{F} \to \overline{\mathbb{R}}$ *de la forma* $\varphi = \sum_{k=1}^{m} b_k I_{E_k}$, *donde* $m \in \mathbb{N}$, $b_1, \ldots, b_m$ *son números reales y* $E_1, \ldots, E_m$ *son conjuntos medibles, es medible.*

Demostración

Sea $E = \cup_{k=1}^{m} E_k$, $T = \{1, \ldots, m\}$ y, si $A = \{i_1, \ldots, i_k\} \subset T$ y $T - A = \{j_1, \ldots, j_{m-k}\}$, definamos:

$$E_A = E \cap E_{i_1} \cap \cdots \cap E_{i_k} \cap E_{j_1}^c \cap \cdots \cap E_{j_{m-k}}^c,$$

$$c_A = \begin{cases} b_{i_1} + \cdots + b_{i_k} & \text{si } A \neq \emptyset \\ 0 & \text{si } A = \emptyset \end{cases}$$

Entonces E_A es medible para cualquier $A \subset T$, $E = \bigcup_{\{A \subset T\}} E_A$, y, si A y B son dos subconjuntos distintos de T, E_A y E_B son ajenos.

Además, si $x \in E_A$ y $A = \{i_1, \ldots, i_k\}$, $\varphi(x) = b_{i_1} + \cdots + b_{i_k} = c_A$, por lo tanto $\varphi = \sum_{A \subset T} c_A I_{E_A}$, de lo cual se sigue inmediatamente que φ es medible.

■

Definición 7.2. *Diremos que una función medible $\varphi : \mathbb{F} \to \overline{\mathbb{R}}$ es simple si tiene la forma $\varphi = \sum_{k=1}^{m} b_k I_{E_k}$, donde $b_1, \ldots, b_m$ son números reales y $E_1, \ldots, E_m$ son conjuntos medibles.*

El resultado siguiente es la base para la definición de la integral de una función medible no negativa y también para demostrar algunas de las propiedades de las funciones medibles, así como de sus integrales.

Teorema 7.3. *Sea $f : \mathbb{F} \to \overline{\mathbb{R}}$ una función medible no negativa, entonces existe una sucesión no decreciente de funciones simples no negativas $\varphi_n : \mathbb{F} \to \mathbb{R}$ tales que $\lim_{n \rightsquigarrow \infty} \varphi_n(x) = f(x)$ para cualquier $x \in \mathbb{F}$.*

Demostración

Para cada $n \in \mathbb{N}$, definamos:

$$\varphi_n(x) = \begin{cases} \sum_{m=1}^{n2^n} \frac{m-1}{2^n} I_{\left\{y \in \mathbb{F} : \frac{m-1}{2^n} \leq f(y) < \frac{m}{2^n}\right\}}(x) & \text{si } f(x) < n \\ n & \text{si } f(x) \geq n \end{cases}$$

Si $f(x) = \infty$, entonces $\varphi_n(x) = n$ para cualquier $n \in \mathbb{N}$, así que $\varphi_n(x) \leq \varphi_{n+1}(x)$ y $\lim_{n \rightsquigarrow \infty} \varphi_n(x) = \infty = f(x)$.

Si $f(x) < n$ para alguna $n \in \mathbb{N}$, sea m el único número natural tal que $\frac{m-1}{2^n} \leq f(x) < \frac{m}{2^n}$. Entonces, como $\varphi_n(x) = \frac{m-1}{2^n}$, se tiene $f(x) - \frac{1}{2^n} < \varphi_n(x) \leq f(x)$, así que $\lim_{n \rightsquigarrow \infty} \varphi_n(x) = f(x)$.

Ahora bien, como $\frac{2(m-1)}{2^{n+1}} \leq f(x) < \frac{2m}{2^{n+1}}$, se tiene que $\frac{2m-2}{2^{n+1}} \leq f(x) < \frac{2m-1}{2^{n+1}}$ o bien $\frac{2m-1}{2^{n+1}} \leq f(x) < \frac{2m}{2^{n+1}}$. En el primer caso, se tiene $\varphi_{n+1}(x) = \frac{2m-2}{2^{n+1}} = \frac{m-1}{2^n} = \varphi_n(x)$ mientras que en el segundo, se tiene $\varphi_{n+1}(x) = \frac{2m-1}{2^{n+1}} > \frac{2m-2}{2^{n+1}} = \varphi_n(x)$. Así que, en cualquier caso, $\varphi_n(x) \leq \varphi_{n+1}(x)$.

Así que, φ_n es una sucesión no decreciente de funciones simples no negativas tales que $\lim_{n\to\infty} \varphi_n(x) = f(x)$ para cualquier $x \in \mathbb{F}$.

■

TEOREMA 7.4. *Sea* $f : \mathbb{F} \to \overline{\mathbb{R}}$ *una función medible, entonces existe una sucesión de funciones simples* φ_n *tales que* $\lim_{n\rightsquigarrow\infty} \varphi_n(x) = f(x)$ *para cualquier* $x \in \mathbb{F}$ *y* $|\varphi_n| \leq |f|$ *para cualquier* $n \in \mathbb{N}$.

Demostración

Se tiene:

$f^+ = \max(f, 0)$,

$f^- = \max(-f, 0)$,

f^+ y f^- son funciones medibles no negativas, las cuales no pueden tomar el valor ∞ en el mismo punto.

Así que si $\{\varphi_n\}_{n\in\mathbb{N}}$ y $\{\psi_n\}_{n\in\mathbb{N}}$ son sucesiones no decrecientes de funciones simples no negativas tales que $\lim_{n\rightsquigarrow\infty} \varphi_n(x) = f^+(x)$ y $\lim_{n\rightsquigarrow\infty} \psi_n(x) = f^-(x)$ para cualquier $x \in \mathbb{F}$, entonces $\varphi_n - \psi_n$ es una función simple para cualquier $n \in \mathbb{N}$ y $\lim_{n\rightsquigarrow\infty} [\varphi_n(x) - \psi_n(x)] = f^+(x) - f^-(x) = f(x)$ para cualquier $x \in \mathbb{F}$. Además, como $\varphi_n \leq f^+$ y $\psi_n \leq f^-$ para cualquier $n \in \mathbb{N}$, entonces $|\varphi_n - \psi_n| \leq \varphi_n + \psi_n = |f|$ para cualquier $n \in \mathbb{N}$.

■

Si $\varphi = \sum_{k=1}^{m} b_k I_{G_k}$ es una función simple entonces el conjunto de los valores que toma es finito. Sea $\{a_1, \ldots, a_n\}$ el conjunto formado por todos los distintos posibles valores no nulos de φ y, para $k \in \{1, \ldots, n\}$, sea $E_k = \{x \in \mathbb{F} : \varphi(x) = a_k\}$, entonces los conjuntos $E_1, \ldots, E_n$ son ajenos por parejas y $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$. Esta última sumatoria será llamada la **representación canónica** de φ.

Ahora demostraremos las propiedades básicas de las funciones medibles con valores en $\overline{\mathbb{R}}$.

PROPOSICIÓN 7.4. *Sea* $g_1 : \mathbb{F} \to \overline{\mathbb{R}}$, $g_2 : \mathbb{F} \to \overline{\mathbb{R}}$, ... *una sucesión de funciones medibles, entonces:*

(i) *Para cualquier* $n \in \mathbb{N}$, *las funciones* $\min\{g_1, \ldots, g_n\}$ *y* $\max\{g_1, \ldots, g_n\}$ *son medibles.*
(ii) *Las funciones* $\inf\{g_1, g_2, \ldots\}$ *y* $\sup\{g_1, g_2, \ldots\}$ *son medibles.*

Demostración

Para cualquier $y \in \mathbb{F}$, se tiene:

$\{x \in \mathbb{F} : \min\{g_1, \ldots, g_n\}(x) \geq y\} = \bigcap_{k=1}^{n} \{x \in \mathbb{F} : g_k(x) \geq y\} \in \Im$,

$\{x \in \mathbb{F} : \max\{g_1, \ldots, g_n\}(x) \leq y\} = \bigcap_{k=1}^{n} \{x \in \mathbb{F} : g_k(x) \leq y\} \in \Im$,

$\{x \in \mathbb{F} : \text{ínf}\{g_1, g_2, \ldots\}(x) \geq y\} = \bigcap_{k=1}^{\infty} \{x \in \mathbb{F} : g_k(x) \geq y\} \in \Im$,

$\{x \in \mathbb{F} : \sup\{g_1, g_2, \ldots\}(x) \leq y\} = \bigcap_{k=1}^{\infty} \{x \in \mathbb{F} : g_k(x) \leq y\} \in \Im$,

de lo cual se sigue el resultado.

■

COROLARIO 7.1. *Sea $f_1 : \mathbb{F} \to \overline{\mathbb{R}}$, $f_2 : \mathbb{F} \to \overline{\mathbb{R}}$, ... una sucesión de funciones medibles, entonces las funciones* $\liminf f_n$ *y* $\limsup f_n$ *son medibles.*

Demostración

La sucesión $g_n = \text{ínf}\{f_j : j \geq n\}$ es no decreciente y:

$\liminf_{n \rightsquigarrow \infty} f_n = \lim_{n \rightsquigarrow \infty} g_n = \sup\{g_n : n \in \mathbb{N}\}$.

Así que $\liminf_{n \rightsquigarrow \infty} f_n$ es medible.

La sucesión $h_n = \sup\{f_j : j \geq n\}$ es no creciente y:

$\limsup f_n = \lim_{n \rightsquigarrow \infty} h_n = \text{ínf}\{h_n : n \in \mathbb{N}\}$.

Así que $\limsup f_n$ es medible.

■

COROLARIO 7.2. *Sea $g_1 : \mathbb{F} \to \overline{\mathbb{R}}$, $g_2 : \mathbb{F} \to \overline{\mathbb{R}}$, ... una sucesión de funciones medibles tales que* $\lim_{n \rightsquigarrow \infty} g_n(x)$ *existe para cualquier $x \in \mathbb{F}$, entonces la función $g : \mathbb{F} \to \overline{\mathbb{R}}$ definida por* $g(x) = \lim_{n \to \infty} g_n(x)$ *es medible.*

LEMA 7.1. *Si una función $f : \mathbb{F} \to \overline{\mathbb{R}}$ es medible, entonces f^+ y f^- son medibles.*

Demostración

La función idénticamente cero es medible, así que entonces $f^+ = \text{máx}\{f, 0\}$, es medible.

Para cualquier $y \in \mathbb{R}$, se tiene $\{x \in \mathbb{F} : [-f](x) \leq y\} = \{x \in \mathbb{F} : f(x) \geq y\}$, así que la función $-f$ es medible. Por lo tanto, $f^- = \text{máx}\{-f, 0\}$ es medible.

■

PROPOSICIÓN 7.5. *Sean $f : \mathbb{F} \to \overline{\mathbb{R}}$ y $g : \mathbb{F} \to \overline{\mathbb{R}}$ dos funciones medibles y $c \in \mathbb{R}$. Entonces las funciones $f + c$, cf y fg son medibles.*

Demostración

Sean φ_n, Φ_n, ϕ_n y Ψ_n sucesiones no decrecientes de funciones simples no negativas tales que $\lim_{n \to \infty} \varphi_n(x) = f^+(x)$, $\lim_{n \rightsquigarrow \infty} \Phi_n(x) = f^-(x)$, $\lim_{n \rightsquigarrow \infty} \phi_n(x) = g^+(x)$ y $\lim_{n \rightsquigarrow \infty} \Psi_n(x) = g^-(x)$, respectivamente, para cualquier $x \in \mathbb{F}$.

Las funciones $\varphi_n - \Phi_n + c$, $c\varphi_n - c\Phi_n$ y $\varphi_n\phi_n + \Phi_n\Psi_n - \varphi_n\Psi_n - \Phi_n\phi_n$ son simples y se tiene:

$\text{lím}_{n\rightsquigarrow\infty}\left[\varphi_n - \Phi_n + c\right](x) = f(x) + c$,

$\left[c\varphi_n - c\Phi_n\right](x) = cf(x)$,

$\text{lím}_{n\rightsquigarrow\infty}\left[\varphi_n\phi_n + \Phi_n\Psi_n - \varphi_n\Psi_n - \Phi_n\phi_n\right](x)$

$= \text{lím}_{n\rightsquigarrow\infty}\left[\varphi_n - \Phi_n\right](x)\left[\phi_n - \Psi_n\right](x) = [fg](x)$.

Así que $f + c$, cf y fg son medibles.

■

El siguiente resultado básicamente expresa que la suma de dos funciones medibles es medible. Sin embargo hay que formularlo bien pues al tratarse de funciones con valores en $\overline{\mathbb{R}}$, la suma de las dos funciones podría no estar definida en algunos puntos.

Proposición 7.6. *Sean $f : \mathbb{F} \to \overline{\mathbb{R}}$ y $g : \mathbb{F} \to \overline{\mathbb{R}}$ dos funciones medibles y $h : \mathbb{F} \to \overline{\mathbb{R}}$ una función tal que $h(x) = f(x) + g(x)$ en todos los puntos $x \in \mathbb{F}$ para los cuales $f(x) + g(x)$ esté definida y h es constante en el conjunto de puntos $x \in \mathbb{F}$ para los cuales $f(x) + g(x)$ no esté definida. Entonces h es medible.*

Demostración

Sean φ_n, Φ_n, ϕ_n y Ψ_n sucesiones no decrecientes de funciones simples no negativas tales que $\text{lím}_{n\to\infty}\varphi_n(x) = f^+(x)$, $\text{lím}_{n\rightsquigarrow\infty}\Phi_n(x) = f^-(x)$, $\text{lím}_{n\rightsquigarrow\infty}\phi_n(x) = g^+(x)$ y $\text{lím}_{n\rightsquigarrow\infty}\Psi_n(x) = g^-(x)$, respectivamente, para cualquier $x \in \mathbb{F}$.

Denotemos por Γ al conjunto de puntos $x \in \mathbb{F}$ para los cuales $f(x) + g(x)$ no está definida.

Γ es medible, las funciones $\varphi_n - \Phi_n + \phi_n - \Psi_n$ son simples y, para cualquier $x \in \Gamma^c$, se tiene:

$\text{lím}_{n\rightsquigarrow\infty}\left[\varphi_n - \Phi_n + \phi_n - \Psi_n\right](x)$

$= \text{lím}_{n\rightsquigarrow\infty}\left[\varphi_n - \Phi_n + \phi_n - \Psi_n\right](x) = f(x) + g(x) = h(x)$.

Así que hI_{Γ^c} es medible.

Sea $\gamma \in \overline{\mathbb{R}}$ el valor constante que toma h en el conjunto de puntos $x \in \mathbb{F}$ para los cuales $f(x) + g(x)$ no está definida.

Si B es un boreliano de $\overline{\mathbb{R}}$, se tiene:

$$h^{-1}(B) = \begin{cases} \left(hI_{\Gamma^c}\right)^{-1}(B) & \text{si } \gamma \notin B \\ \left[\left(hI_{\Gamma^c}\right)^{-1}\right] \cup \Gamma & \text{si } \gamma \in B \end{cases}$$

Así que h es medible.

■

Recordemos que en el capítulo 2 establecimos la convención de considerar a la suma de dos funciones $f : \mathbb{F} \to \overline{\mathbb{R}}$ y $g : \mathbb{F} \to \overline{\mathbb{R}}$ como la función $h : \mathbb{F} \to \overline{\mathbb{R}}$ definida de la siguiente manera:

$$h(x) = \begin{cases} f(x) + g(x) & \text{si } f(x) + g(x) \text{ está definida} \\ 1 & \text{si } f(x) + g(x) \text{ no está definida} \end{cases}$$

Así que, de acuerdo con la proposición anterior se tiene el siguiente resultado:

COROLARIO 7.3. *Si $f : \mathbb{F} \to \overline{\mathbb{R}}$ y $g : \mathbb{F} \to \overline{\mathbb{R}}$ son dos funciones medibles, entonces $f + g$ es medible.*

TEOREMA 7.5. *Supongamos que el espacio de medida $(\mathbb{F}, \Im, \mu)$ es completo y sean $f : \mathbb{F} \to \overline{\mathbb{R}}$ una función medible y $g : \mathbb{F} \to \overline{\mathbb{R}}$ una función tal que $g = f$ excepto a lo más en un conjunto de medida cero, entonces g es medible.*

Demostración

Sea $E = \{x \in \mathbb{F} : g(x) = f(x)\}$, entonces, para cualquier $y \in \mathbb{R}$, se tiene.

$\{x \in \mathbb{F} : g(x) \leq y\}$

$= (\{x \in \mathbb{F} : g(x) \leq y\} \cap E) \cup (\{x \in \mathbb{F} : g(x) \leq y\} \cap E^c)$

$= (\{x \in \mathbb{F} : f(x) \leq y\} \cap E) \cup \{x \in E^c : g(x) \leq y\}.$

El conjunto $\{x \in E^c : g(x) \leq y\}$ está contenido en un conjunto de medida cero, por lo tanto es medible. Así que $\{x \in \mathbb{F} : g(x) \leq y\}$ es medible.

■

En general, cuando se tiene un espacio de medida $(\mathbb{F}, \Im, \mu)$, los conjuntos de medida cero son considerados pequeños al grado de que pueden ser despreciados. De esta forma, dos funciones medibles que sean iguales excepto en un conjunto de medida cero son esencialmente la misma función. Esta idea puede formalizarse definiendo una relación de equivalencia dentro del conjunto de las funciones medibles:

DEFINICIÓN 7.3. *Diremos que dos funciones medibles son equivalentes si el conjunto de puntos donde son distintas tiene medida μ cero.*

Se verifica inmediatamente que la relación así definida es efectivamente una relación de equivalencia, de manera que, mediante ella, el conjunto de las funciones medibles queda partido en clases de equivalencia, cada una de las cuales está formada por funciones medibles que son iguales excepto en un conjunto de medida cero.

Si f es una función medible, denotaremos por $[f]$ a la clase equivalencia que contiene a f.

PROPOSICIÓN 7.7. *Sea $(\mathbb{F}, \Im_0, \mu)$ un espacio de medida y $\Im$ la completación de $\Im_0$ con respecto a μ. Si $f : \mathbb{F} \mapsto \overline{\mathbb{R}}$ es una función $\Im$-medible no negativa, entonces existe un conjunto $B \in \Im_0$ de medida cero y una función $\psi : \mathbb{F} \mapsto \overline{\mathbb{R}}$, $\Im_0$-medible no negativa, tal que $f = \psi I_{B^c} + f I_B$.*

Demostración

Para cada $n \in \mathbb{N}$, sea:

$$\varphi_n(x) = \begin{cases} \sum_{m=0}^{n2^n-1} \frac{m}{2^n} I_{\left\{y \in \mathbb{R}: \frac{m}{2^n} \leq f(y) < \frac{m+1}{2^n}\right\}}(x) & \text{si } f(x) < n \\ n & \text{si } f(x) \geq n \end{cases}$$

$A_n^{n2^n} = \{y \in \mathbb{F} : f(y) \geq n\}$,

$A_n^m = \left\{y \in \mathbb{F} : \frac{m}{2^n} \leq f(y) < \frac{m+1}{2^n}\right\}$, para $m \in \{0, \ldots, n2^n - 1\}$.

Entonces:

$\bigcup_{m=0}^{n2^n-1} A_n^m = \{y \in \mathbb{F} : f(y) < n\}$.

Para $n \in \mathbb{N}$ y $m \in \{0, \ldots, n2^n\}$, sean $B_n^m \in \Im_0$ y C_n^m un conjunto de medida cero tales que $A_n^m = B_n^m \cup C_n^m$ y $B_n^m \cap C_n^m = \emptyset$. Entonces existe un conjunto $B \in \Im_0$ de medida cero tal que $\bigcup_{n=1}^{\infty} \bigcup_{m=0}^{n2^n} C_n^m \subset B$ y, para cada $n \in \mathbb{N}$, la función $\psi_n = \sum_{m=0}^{n2^n-1} \frac{m}{2^n} I_{B_n^m} + n I_{B_n^{n2^n}}$ es $\Im_0$-medible y:

$\varphi_n = \sum_{m=0}^{n2^n-1} \frac{m-1}{2^n} I_{A_n^m} + n I_{A_n^{n2^n}}$

$= \sum_{m=0}^{n2^n-1} \frac{m-1}{2^n} I_{B_n^m} + n I_{B_n^{n2^n}} + \sum_{m=0}^{n2^n-1} \frac{m-1}{2^n} I_{C_n^m} + n I_{C_n^{n2^n}}$.

Para cualquier $n \in \mathbb{N}$, si $y \in B^c$, entonces $y \in B_n^m$ para alguna $m \in \{0, \ldots, n2^n\}$. Por lo tanto, $\psi_n(y) = \varphi_n(y)$.

Así que $\psi_n I_{B^c} = \varphi_n I_{B^c}$ y entonces ψ_n converge sobre B^c.

Sea:

$$\psi(y) = \begin{cases} \lim_{n \rightsquigarrow \infty} \psi_n(y) & \text{si } y \in B^c \\ 0 & \text{si } y \in B \end{cases}$$

Entonces ψ es $\Im_0$-medible y $\psi = f$ sobre B^c. Por lo tanto:

$f = f I_{B^c} + f I_B = \psi I_{B^c} + f I_B$.

■

7.4. Funciones medibles con valores en $\overline{\mathbb{R}}^n$

La medibilidad de una función $f : \mathbb{F} \to \mathbb{R}^n$ (resp. $f : \mathbb{F} \to \overline{\mathbb{R}}^n$) será entendida considerando sobre $\mathbb{R}^n$ (resp. $\overline{\mathbb{R}}^n$) la σ-álgebra de los conjuntos borelianos en $\mathbb{R}^n$ (resp. $\overline{\mathbb{R}}^n$).

Dada una función $f : \mathbb{F} \to \mathbb{R}^n$, podemos definir, para cada $k \in \{1, 2, \ldots, n\}$, una función $f_k : \mathbb{F} \to \mathbb{R}$ la cual asocia a cada $x \in \mathbb{F}$, la k-ésima coordenada de la imagen de x bajo f, de manera que f se puede escribir como $(f_1, f_2, \ldots, f_n)$. Las funciones $f_1, f_2, \ldots, f_n$ definidas de esta manera será llamadas las componentes de la función f.

Obviamente, si partimos de n funciones cualesquiera, $f_1, f_2, \ldots, f_n$, de $\mathbb{F}$ en $\mathbb{R}$, podemos definir la función $(f_1, f_2, \ldots, f_n) : \mathbb{F} \to \mathbb{R}^n$ mediante la relación:

$(f_1, f_2, \ldots, f_n)(x) = (f_1(x), f_2(x), \ldots, f_n(x))$.

Las componentes de esta función así definida son $f_1, f_2, \ldots, f_n$.

De manera similar, podemos definir las componentes de una función $f : \mathbb{F} \to \overline{\mathbb{R}}^n$.

PROPOSICIÓN 7.8. *Sea $f : \mathbb{F} \to \mathbb{R}^n$ y $f_1, f_2, \ldots, f_n$ las componentes de f. Entonces f es medible si y sólo si se cumple cualquiera de las siguientes propiedades:*

(i) $\bigcap_{k=1}^{n} \{x \in \mathbb{F} : f_k(x) \leq y_k]\} \in \Im$ *para cualquier vector* $(y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$.
(ii) $\bigcap_{k=1}^{n} \{x \in \mathbb{F} : f_k(x) < y_k]\} \in \Im$ *para cualquier vector* $(y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$.
(iii) $\bigcap_{k=1}^{n} \{x \in \mathbb{F} : f(x) \in (a_k, b_k]\} \in \Im$
para cualesquiera $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$.
(iv) $\bigcap_{k=1}^{n} \{x \in \mathbb{F} : f(x) \in [a_k, b_k)\} \in \Im$
para cualesquiera $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$.
(v) $\bigcap_{k=1}^{n} \{x \in \mathbb{F} : f(x) \in (a_k, b_k)\} \in \Im$
para cualesquiera $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$.
(vi) $\bigcap_{k=1}^{n} \{x \in \mathbb{F} : f(x) \in [a_k, b_k]\} \in \Im$
para cualesquiera $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$.

TEOREMA 7.6. *Sea $f : \mathbb{F} \to \mathbb{R}^n$ y $f_1, f_2, \ldots, f_n$ las componentes de f. Entonces f es medible si y sólo si $f_1, f_2, \ldots, f_n$ son medibles.*

Demostración

Supongamos que $f_1, f_2, \ldots, f_n$ son medibles, entonces, para cualquier $k \in \{1, 2, \ldots, n\}$, $\{x \in \mathbb{F} : f_k(x) \leq y_k]\} \in \Im$ para cualquier $y_k \in \mathbb{R}$. Por lo tanto:

$\bigcap_{k=1}^{n} \{x \in \mathbb{F} : f_k(x) \leq y_k]\} \in \Im$ para cualquier vector $(y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$.

Inversamente, supongamos que f es medible y sean $k \in \{1, 2, \ldots, n\}$ y $y_k \in \mathbb{R}$. Para cada $j \in \{1, 2, \ldots, n\}$, definamos:

$$I_j = \begin{cases} (-\infty, y_k] & \text{si } j = k \\ \mathbb{R} & \text{si } j \neq k \end{cases}$$

Entonces:

$\{x \in \mathbb{F} : f_k(x) \leq y_k]\} = \bigcap_{j=1}^{n} \{x \in \mathbb{F} : f_j(x) \in I_j\} \in \Im$.

■

PROPOSICIÓN 7.9. *Sea $f : \mathbb{F} \to \overline{\mathbb{R}}^n$ y $f_1, f_2, \ldots, f_n$ las componentes de f. Entonces f es medible si y sólo si:*

$\bigcap_{k=1}^{n} \{x \in \mathbb{F} : f_k(x) \leq y_k]\} \in \Im$

para cualquier vector $(y_1, y_2, \ldots, y_n) \in \overline{\mathbb{R}}^n$.

TEOREMA 7.7. *Sea* $f : \mathbb{F} \to \overline{\mathbb{R}}^n$ *y* $f_1, f_2, \ldots, f_n$ *las componentes de* f*. Entonces* f *es medible si y sólo si* $f_1, f_2, \ldots, f_n$ *son medibles.*

Demostración

Supongamos que $f_1, f_2, \ldots, f_n$ son medibles, entonces, para cualquier $k \in \{1, 2, \ldots, n\}$, $\{x \in \mathbb{F} : f_k(x) \leq y_k]\} \in \Im$ para cualquier $y_k \in \overline{\mathbb{R}}$. Por lo tanto:

$\bigcap_{k=1}^{n} \{x \in \mathbb{F} : f_k(x) \leq y_k]\} \in \Im$ para cualquier vector $(y_1, y_2, \ldots, y_n) \in \overline{\mathbb{R}}^n$.

Inversamente, supongamos que f es medible y sean $k \in \{1, 2, \ldots, n\}$ y $y_k \in \mathbb{R}$. Para cada $j \in \{1, 2, \ldots, n\}$, definamos:

$$I_j = \begin{cases} [-\infty, y_k] & \text{si } j = k \\ \overline{\mathbb{R}} & \text{si } j \neq k \end{cases}$$

Entonces:

$\{x \in \mathbb{F} : f_k(x) \leq y_k]\} = \bigcap_{j=1}^{n} \{x \in \mathbb{F} : f_j(x) \in I_j\} \in \Im$.

■

TEOREMA 7.8. *Toda función continua* $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ *es medible.*

Demostración

Sea $\mathcal{H} = \{B \subset \mathbb{R}^n : f^{-1}(B) \in \Im\}$.

$\mathcal{H}$ es una σ-álgebra de subconjuntos de $\mathbb{F}$ la cual contiene a los conjuntos abiertos ya que f es continua. Por lo tanto, $\mathcal{H}$ contiene a la σ-álgebra generada por los conjuntos abiertos de $\mathbb{R}^n$, es decir, contiene a los borelianos de $\mathbb{R}^n$ ya que $\mathfrak{B}(\mathbb{R}^n)$ está generada por los conjuntos abiertos.

Por lo tanto, f es medible.

■

COROLARIO 7.4. *Sea* $f : \mathbb{F} \to \mathbb{R}$ *una función medible, entonces las funciones* $g = \frac{1}{f} I_{\{x \in \mathbb{F} : f(x) \neq 0\}}$ *y* $h = |f|^{\alpha}$*, donde* $\alpha \geq 0$*, son medibles.*

El siguiente resultado permite realizar el vínculo entre la teoría de la medida y la teoría de integración con un enfoque geométrico. Cuando uno quiere calcular el área de la región comprendida bajo la gráfica de una función no negativa, se integra tal función . En otras palabras, la integral de Riemann de una función no negativa corresponde al área bajo la gráfica de esa función. Lo mismo ocurre con la integral definida por H. Lebesgue. Para que este resultado adquiera sentido se requiere que, dada una función boreliana no negativa, la región comprendida bajo la gráfica de la función sea un conjunto al cual se le puede asignar área, es decir, que sea Lebesgue medible. Como se muestra a continuación, tal región es un conjunto boreliano de $\mathbb{R}^2$.

PROPOSICIÓN 7.10. *Sea $g : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ cualquier función boreliana no negativa. Entonces la función $h : \mathbb{R}^{n+1} \mapsto \mathbb{R}$ definida por $h(x, y) = I_{[0,g(x))}(y)$, donde $x \in \mathbb{R}^n$ y $y \in \mathbb{R}$, es boreliana.*

Demostración

Para cada $m \in \mathbb{N}$, sea:

$$\varphi_m(x) = \begin{cases} \sum_{k=0}^{m2^m} \frac{k-1}{2^m} I_{\left[\frac{k-1}{2^m} \leq g < \frac{k}{2^m}\right]} & \text{si } g(x) < m \\ \infty & \text{si } g(x) \geq m \end{cases}$$

De acuerdo con la demostración de la proposición anterior, φ_m es una sucesión no decreciente de funciones no negativas tales que $\lim_{m \rightsquigarrow \infty} \varphi_m(x) = g(x)$ para cualquier $x \in \mathbb{R}^n$. Por lo tanto:

$$\{(x, y) \in \mathbb{R}^{n+1} : 0 \leq y < g(x)\} = \cup_{m=1}^{\infty} \{(x, y) \in \mathbb{R}^{n+1} : 0 \leq y < \varphi_m(x)\}$$

$$= \cup_{m=1}^{\infty} \left(\cup_{k=0}^{m2^m} \left[\left[\frac{k-1}{2^m} \leq g < \frac{k}{2^m} \right] \times \left[0, \frac{k-1}{2^m}\right) \right] \right).$$

Así que el conjunto $\{(x, y) \in \mathbb{R}^{n+1} : h(x, y) = 1\}$ es boreliano.

■

COROLARIO 7.5. *Sea $g : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ una función boreliana no negativa. Entonces los siguientes conjuntos son borelianos de $\mathbb{R}^{n+1}$:*

(i) $\{(x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1} : 0 < y < g(x_1, \ldots, x_n)\}$,
(ii) $\{(x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1} : 0 \leq y < g(x_1, \ldots, x_n)\}$,
(iii) $\{(x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1} : 0 < y \leq g(x_1, \ldots, x_n)\}$,
(iv) $\{(x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1} : 0 \leq y \leq g(x_1, \ldots, x_n)\}$,
(v) $\{(x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1} : g(x) > 0 \text{ y } 0 \leq y \leq g(x_1, \ldots, x_n)\}$.

7.5. La integral de funciones medibles simples no negativas

En seguida se desarrolla la formulación moderna de la teoría de integración de Lebesgue para el caso de un espacio de medida $(\mathbb{F}, \Im, \mu)$ cualquiera. La idea es definir primero la integral para las funciones simples no negativas, después para cualquier función medible no negativa y finalmente para cualquier función medible.

Como una función simple puede tener varias representaciones, es necesario hacer ver que se puede obtener la integral de la función teniendo cualquiera de sus representaciones.

DEFINICIÓN 7.4. *Si φ es una función simple no negativa con representación canónica $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$, se define la integral de φ, $\int_{\mathbb{F}} \varphi d\mu$, de la siguiente manera:*

$\int_{\mathbb{F}} \varphi d\mu = \sum_{k=1}^{n} a_k m(E_k)$.

Lema 7.2. *Sea* $\varphi = \sum_{j=1}^{m} b_j I_{F_j}$ *una función simple no negativa, donde los conjuntos* $F_1, \ldots, F_m$ *son ajenos por parejas, con representación canónica* $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$*, entonces* $\sum_{j=1}^{m} b_j m(F_j) = \sum_{k=1}^{n} a_k m(E_k)$.

Demostración

Para $k \in \{1, \ldots, n\}$, se tiene $E_k = \cup_{\{j\in\{1,\ldots,m\}:b_j=a_k\}} F_j$, así que:

$\sum_{k=1}^{n} a_k m(E_k) = \sum_{k=1}^{n} a_k \sum_{\{j\in\{1,\ldots,m\}:b_j=a_k\}} m(F_j)$

$= \sum_{k=1}^{n} \sum_{\{j\in\{1,\ldots,m\}:b_j=a_k\}} b_j m(F_j) = \sum_{j=1}^{m} b_j m(F_j)$.

■

Proposición 7.11. *Sea* $\varphi = \sum_{j=1}^{m} b_j I_{F_j}$ *una función simple, donde los coeficientes* $b_1, \ldots, b_m$ *son no negativos, con representación canónica* $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$*, entonces* $\sum_{j=1}^{m} b_j m(F_j) = \sum_{k=1}^{n} a_k m(E_k)$.

Demostración

Los términos de la sumatoria $\sum_{j=1}^{m} b_j I_{F_j}$ en los cuales $b_j = 0$ pueden eliminarse, así que podemos asumir que $b_1, \ldots, b_m$ son números reales positivos.

Sea $F = \cup_{j=1}^{m} F_m$, $T = \{1, \ldots, m\}$ y, si $A = \{i_1, \ldots, i_k\} \subset T$ y $T - A = \{j_1, \ldots, j_{m-k}\}$, definamos:

$F_A = F \cap F_{i_1} \cap \cdots \cap F_{i_k} \cap F_{j_1}^c \cap \cdots \cap F_{j_{m-k}}^c$,

$c_A = b_{i_1} + \cdots + b_{i_k}$.

Entonces $F = \cup_{A\subset T} F_A$, $F_i = \cup_{\{A\subset T: i\in A\}} F_A$ y, si A y B son dos subconjuntos distintos de T, F_A y F_B son ajenos. Por lo tanto:

$m(F_j) = \sum_{\{A\subset T: j\in A\}} m(F_A)$.

Así que:

$\sum_{j=1}^{m} b_j m(F_j) = \sum_{j=1}^{m} b_j \sum_{\{A\subset T: j\in A\}} m(F_A)$

$= \sum_{j=1}^{m} \sum_{\{A\subset T: j\in A\}} b_j m(F_A) = \sum_{A\subset T} \sum_{j\in A} b_j m(F_A)$

$= \sum_{A\subset T} c_A m(F_A)$.

Además, si $x \in F_A$ y $A = \{i_1, \ldots, i_k\}$, $\varphi(x) = b_{i_1} + \cdots + b_{i_k} = c_A$, por lo tanto $\varphi = \sum_{A\subset T} c_A I_{F_A}$.

Así que se tiene:

$\sum_{j=1}^{m} b_j I_{F_j} = \sum_{A\subset T} c_A I_{F_A}$,

$\sum_{j=1}^{m} b_j m(F_j) = \sum_{A\subset T} c_A m(F_A)$.

Pero como $\sum_{i=1}^{n} a_i I_{E_i} = \sum_{j=1}^{m} b_j I_{F_j}$, se tiene $\sum_{i=1}^{n} a_i I_{E_i} = \sum_{A\subset T} c_A I_{E_A}$, así que, por el lema anterior:

$\sum_{j=1}^{m} b_j m(F_j) = \sum_{A\subset T} c_A m(F_A) = \sum_{k=1}^{n} a_k m(E_k)$.

■

COROLARIO 7.6. *Sea $\varphi = \sum_{j=1}^{m} b_j I_{F_j}$ una función simple, donde los coeficientes $b_1, \ldots, b_m$ son no negativos, entonces:*

$\int_{\mathbb{F}} \varphi d\mu = \sum_{j=1}^{m} b_j m(F_j)$.

PROPOSICIÓN 7.12. *Sean φ y ψ dos funciones simples no negativas, entonces:*

(i) $\int_{\mathbb{F}} [a\varphi + b\psi]\, d\mu = a\int_{\mathbb{F}} \varphi d\mu + b\int_{\mathbb{F}} \psi d\mu$ *para cualesquiera números reales a y b no negativos.*
(ii) *Si $\varphi \le \psi$, entonces $\int_{\mathbb{F}} \varphi d\mu \le \int_{\mathbb{F}} \psi d\mu$.*

Demostración

i. Sean $\sum_{k=1}^{n} a_k I_{E_k}$ y $\sum_{k=1}^{m} b_k I_{F_k}$ las representaciones canónicas de φ y ψ, respectivamente. Entonces:

$a\varphi + b\psi = \sum_{k=1}^{n} a a_k I_{E_k} + \sum_{k=1}^{m} b b_k I_{F_k}..$

Así que:

$\int_{\mathbb{F}} [a\varphi + b\psi]\, d\mu = \sum_{k=1}^{n} a a_k m(E_k) + \sum_{k=1}^{m} b b_k m(F_k)$

$= a\int_{\mathbb{F}} \varphi d\mu + b\int_{\mathbb{F}} \psi d\mu$.

ii. Si $\varphi \le \psi$, entonces $\psi - \varphi$ es una función simple no negativa y $\psi = \varphi + (\psi - \varphi)$, así que:

$\int_{\mathbb{F}} \psi d\mu = \int_{\mathbb{F}} \varphi d\mu + \int_{\mathbb{F}} [\psi - \varphi]\, d\mu$.

Por lo tanto:

$\int_{\mathbb{F}} \psi d\mu - \int_{\mathbb{F}} \varphi d\mu = \int_{\mathbb{F}} [\psi - \varphi]\, d\mu \ge 0$.

■

TEOREMA 7.9. *Sea φ una función simple no negativa. Entonces, la función $m : \Im \to \mathbb{R}$, definida por $m(E) = \int_E \varphi d\mu$, es una medida.*

Demostración

Obviamente, m es no negativa y, por la proposición anterior, es finitamente aditiva.

Sea $\sum_{k=1}^{n} a_k I_{E_k}$ la representación canónica de φ, A_n una sucesión creciente de conjuntos medibles y $A = \cup_{n=1}^{\infty} A_n$. Entonces:

$\int_{A_n} \varphi d\mu = \sum_{k=1}^{n} a_k m\left(A_n \cap E_k\right)$,

$\int_{A} \varphi d\mu = \sum_{k=1}^{n} a_k m\left(A \cap E_k\right)$.

Así que:

$\lim_{n \rightsquigarrow \infty} m(A_n) = \lim_{n \to \infty} \int_{A_n} \varphi d\mu = \lim_{n \to \infty} \sum_{k=1}^{n} a_k m\left(A_n \cap E_k\right)$

$= \sum_{k=1}^{n} a_k \lim_{n \rightsquigarrow \infty} m\left(A_n \cap E_k\right) = \sum_{k=1}^{n} a_k m\left(A \cap E_k\right)$

$= \int_{A} \varphi d\mu = m(A)$.

■

7.6. La integral de funciones medibles no negativas

Para definir la integral de una función medible no negativa podríamos utilizar el hecho de que se puede aproximar, por abajo, mediante una sucesión no decreciente de funciones simples no negativas y definiendo entonces la integral de la función como el límite de las integrales de las funciones simples que aproximan a la función. Con este método, sería necesario demostrar que el valor que se obtiene es el mismo para cualquier sucesión de funciones simples no negativas cuyo límite sea la función medible dada. La siguiente definición evita tener que hacer eso y es más cómoda de trabajar.

Definición 7.5. *Si f es una función medible no negativa, se define la integral de f, $\int_{\mathbb{F}} f d\mu$, de la siguiente manera:*

$\int_{\mathbb{F}} f d\mu = \sup\left\{ \int_{\mathbb{F}} \varphi d\mu : \varphi \text{ es simple y } 0 \leq \varphi \leq f \ \right\}$.

Definición 7.6. *Si f es una función medible no negativa y E es un conjunto medible, se define:*

$\int_{E} f d\mu = \int_{\mathbb{F}} I_E f d\mu$.

Podemos demostrar inmediatamente el primero de los teoremas de convergencia de la integral, los cuales muestran claramente, para el caso de funciones definidas sobre $\mathbb{R}$, la superioridad de la integral de Lebesgue con respecto a la integral de Riemann

Teorema 7.10 (**Teorema de la convergencia monótona**). *Sea f_n una sucesión no decreciente de funciones medibles no negativas, entonces:*

$\int_{\mathbb{F}} \lim_{n \rightsquigarrow \infty} f_n d\mu = \lim_{n \to \infty} \int_{\mathbb{F}} f_n d\mu$.

Demostración

Sea $f = \text{lím}_{n \rightsquigarrow \infty} f_n$ y φ una función simple no negativa tal que $\varphi \leq f$, $\alpha \in (0,1)$ y $A_n = \{x \in \mathbb{F}: f_n(x) \geq \alpha\varphi(x)\}$. Entonces, la sucesión A_n es creciente y $\cup A_n = \mathbb{F}$. Además, la función $m : \Im \to \mathbb{R}$, definida por $m(E) = \int_E \varphi d\mu$, es una medida, así que:

$\text{lím}_{n \rightsquigarrow \infty} \int_{A_n} \varphi d\mu = \int_{\mathbb{F}} \varphi d\mu$.

Por otra parte, $\alpha \int_{A_n} \varphi d\mu \leq \int_{A_n} f_n d\mu \leq \int_{\mathbb{F}} f_n d\mu$ para cualquier $n \in \mathbb{N}$, así que:

$\alpha \int_{\mathbb{F}} \varphi d\mu = \alpha \, \text{lím}_{n \to \infty} \int_{A_n} \varphi d\mu \leq \text{lím}_{n \rightsquigarrow \infty} \int_{\mathbb{F}} f_n d\mu$.

Haciendo tender α a 1, se obtiene entonces:

$\int_{\mathbb{F}} \varphi d\mu \leq \text{lím}_{n \rightsquigarrow \infty} \int_{\mathbb{F}} f_n d\mu$.

Por lo tanto:

$\int_{\mathbb{F}} f d\mu \leq \text{lím}_{n \rightsquigarrow \infty} \int_{\mathbb{F}} f_n d\mu$.

■

PROPOSICIÓN 7.13. *Sean f y g dos funciones medibles no negativas, entonces:*

(i) $\int_{\mathbb{F}} [af + bg] \, d\mu = a \int_{\mathbb{F}} f d\mu + b \int_{\mathbb{F}} g d\mu$ *para cualesquiera números reales a y b no negativos.*
(ii) *Si $f \leq g$ entonces $\int_{\mathbb{F}} f d\mu \leq \int_{\mathbb{F}} g d\mu$.*
(iii) *Si $f \leq g$ sobre un conjunto $E \in \Im$, entonces $\int_E f d\mu \leq \int_E g d\mu$.*
(iv) *$\int_{\mathbb{F}} f d\mu = 0$ si y sólo si $\mu \{x \in \mathbb{F} : f(x) > 0\} = 0$.*

Demostración

Para la primera, sean φ_n y ψ_n dos sucesiones no decrecientes de funciones simples no negativas tales que $\text{lím}_{n \to \infty} \varphi_n(x) = f(x)$ y $\text{lím}_{n \rightsquigarrow \infty} \psi_n(x) = g(x)$ para cualquier $x \in \mathbb{F}$.

Para cada $n \in \mathbb{N}$, se tiene:

$\int_{\mathbb{F}} [a\varphi_n + b\psi_n] \, d\mu = a \int_{\mathbb{F}} \varphi_n d\mu + b \int_{\mathbb{F}} \psi_n d\mu$.

Así que, por el teorema de la convergencia monótona, se tiene:

$\int_{\mathbb{F}} [af + bg] \, d\mu = a \int_{\mathbb{F}} f d\mu + b \int_{\mathbb{F}} f_n d\mu$.

La segunda propiedad es inmediata de la definición.

Para la tercera propiedad, si $f \leq g$ sobre un conjunto $E \in \Im$, entonces $fI_E \leq gI_E$, así que el resultado se sigue de la segunda propiedad.

Para la cuarta propiedad supongamos que $\int_{\mathbb{F}} f d\mu = 0$.

Para cada $n \in \mathbb{N}$, sea $E_n = \left\{x \in \mathbb{F} : f(x) > \frac{1}{n}\right\}$, entonces, para cualquier $n \in \mathbb{N}$, se tiene:

$\int_{\mathbb{F}} f d\mu \geq \int_{E_n} f d\mu \geq \frac{1}{n}\mu(E_n)$.

Si $\mu(E_n)$ fuera positiva, se tendría $\int_{\mathbb{F}} f d\mu > 0$, por lo tanto $\mu(E_n) = 0$.

Finalmente, $\mu\{x \in \mathbb{F} : f(x) > 0\} = \mu(\cup_{n=1}^{\infty} E_n) = \text{lím}_{n\rightsquigarrow\infty} \mu(E_n) = 0$.

Inversamente, supongamos que $\mu\{x \in \mathbb{F} : f(x) > 0\} = 0$ y sea φ una función medible simple tal que $0 \leq \varphi \leq f$, entonces $\varphi = 0$ casi en todas partes, así que $\int_{\mathbb{F}} \varphi d\mu = 0$; por lo tanto, $\int_{\mathbb{F}} f d\mu = 0$.

■

Teorema 7.11. *Sea f una función medible no negativa. Entonces, la función $m : \Im \to \overline{\mathbb{R}}$, definida por $m(E) = \int_E f d\mu$, es una medida. Además, si h es una función medible no negativa, entonces:*

$\int_{\mathbb{F}} h dm = \int_{\mathbb{F}} h f d\mu$.

Demostración

Obviamente, m es no negativa y, por el inciso 1 de la proposición anterior, es finitamente aditiva.

Sea A_n una sucesión monótona no decreciente de conjuntos medibles y $A = \cup_{n=1}^{\infty} A_n$. Entonces, la sucesión de funciones $(I_{A_n} f)_{n\in\mathbb{N}}$ es no decreciente y $\text{lím}_{n\rightsquigarrow\infty} I_{A_n} = I_A f$, así que, por el teorema de la convergencfia monótona:

$m(A) = \int_A f d\mu = \int_{\mathbb{F}} I_A f d\mu = \text{lím}_{n\to\infty} \int_{\mathbb{F}} I_{A_n} f d\mu = \text{lím}_{n\to\infty} \int_{A_n} f d\mu = \text{lím}_{n\rightsquigarrow\infty} m(A_n)$.

Si $\varphi := \sum_{j=1}^{n} b_j I_{F_j}$ es una función simple no negativa, entonces:

$\int_{\mathbb{F}} \varphi dm = \sum_{j=1}^{n} b_j m\left(I_{F_j}\right) = \sum_{j=1}^{n} b_j \int_{F_j} f d\mu = \int_{F_j} \left(\sum_{j=1}^{n} I_{F_j}\right) f d\mu = \int_{\mathbb{F}} \varphi f d\mu$.

Sea $(\varphi_n)_{n\in\mathbb{N}}$ una sucesión no decreciente de funciones simples no negativas $\varphi_n : \mathbb{F} \to \mathbb{R}$ tales que $\text{lím}_{n\rightsquigarrow\infty} \varphi_n(x) = h(x)$ para cualquier $x \in \mathbb{F}$. Entonces, aplicando el teorema de la convergencia monótona, se tiene:

$\int_{\mathbb{F}} h dm = \text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} \varphi_n dm = \text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} \varphi_n f d\mu = \int_{\mathbb{F}} \text{lím}_{n\rightsquigarrow\infty} \varphi_n f d\mu = \int_{\mathbb{F}} h f d\mu$.

■

Teorema 7.12 (**Lema de Fatou**). *Sea $\{f_n\}_{n\in N}$ una sucesión de funciones medibles no negativas. Entonces:*

$\int_{\mathbb{F}} \text{lím inf}_{n\rightsquigarrow\infty} f_n d\mu \leq \text{lím inf}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} f_n d\mu$

Demostración

La sucesión $g_n = \text{ínf}\{f_j : j \geq n\}$ es no decreciente y $\text{lím ínf}_{n\rightsquigarrow\infty} f_n = \text{lím}_{n\to\infty} g_n$, así que, por el teorema de la convergencia monótona:

$\int_{\mathbb{F}} \text{lím ínf}_{n\rightsquigarrow\infty} f_n d\mu = \text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} g_n d\mu.$

Por otra parte, $g_n \leq f_j$ para cualquier $j \geq n$, así que:

$\int_{\mathbb{F}} g_n d\mu \leq \text{ínf}\left\{\int_{\mathbb{F}} f_j d\mu : j \geq n\right\}.$

Por lo tanto:

$\int_{\mathbb{F}} \text{lím ínf}_{n\rightsquigarrow\infty} f_n d\mu = \text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} g_n d\mu$

$\leq \text{lím}_{n\rightsquigarrow\infty} \text{ínf}\left\{\int_{\mathbb{F}} f_j d\mu : j \geq n\right\} = \text{lím ínf}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} f_n d\mu.$

■

Llamaremos sucesión doble de números reales a cualquier función $x : \mathbb{N} \times \mathbb{N} \to R$. Una sucesión de este tipo será denotada por $(x_{nm})_{n,m\in\mathbb{N}}$, donde $x_{nm} = x(n,m)$.

Diremos que una sucesión doble, $(x_{nm})_{n,m\in\mathbb{N}}$, converge si existe $L \in \mathbb{R}$ tal que, para cualquier $\varepsilon > 0$ existe $N \in \mathbb{N}$ tal que $|x_{nm} - L| < \varepsilon$ para toda pareja de números naturales n y m mayores o iguales a N.

Sea $(x_{nm})_{n,m\in\mathbb{N}}$ una sucesión doble. Considerando esta sucesión en $\overline{\mathbb{R}}$, independientemente de su convergencia o no convergencia, fijando $n \in \mathbb{N}$, la sucesión $(\sup\{x_{ij} : i \geq n, j \geq m\})_{m\in\mathbb{N}}$ es no creciente, así que el límite $\text{lím}_{m\rightsquigarrow\infty} \sup\{x_{ij} : i \geq n, j \geq m\}$ existe. Además, si $n_1, n_2 \in \mathbb{N}$ y $n_1 < n_2$, entonces $\sup\{x_{ij} : i \geq n_1, j \geq m\} \geq \sup\{x_{ij} : i \geq n_2, j \geq m\}$, así que:

$\text{lím}_{m\rightsquigarrow\infty} \sup\{x_{ij} : i \geq n_1, j \geq m\} \geq \text{lím}_{m\rightsquigarrow\infty} \sup\{x_{ij} : i \geq n_2, j \geq m\}.$

Por lo tanto, la sucesión $(\text{lím}_{m\rightsquigarrow\infty} \sup\{x_{ij} : i \geq n, j \geq m\})_{n\in\mathbb{N}}$ es no creciente, así que el límite $\text{lím}_{n\to\infty} \text{lím}_{m\rightsquigarrow\infty} \sup\{x_{ij} : i \geq n, j \geq m\}$ existe.

De la misma manera, fijando $n \in \mathbb{N}$, la sucesión $(\text{ínf}\{x_{ij} : i \geq n, j \geq m\})_{m\in\mathbb{N}}$ es no decreciente, así que el límite $\text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{x_{ij} : i \geq n, j \geq m\}$ existe. Además, si $n_1, n_2 \in \mathbb{N}$ y $n_1 < n_2$, entonces $\text{ínf}\{x_{ij} : i \geq n_1, j \geq m\} \leq \text{ínf}\{x_{ij} : i \geq n_2, j \geq m\}$, así que:

$\text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{x_{ij} : i \geq n_1, j \geq m\} \leq \text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{x_{ij} : i \geq n_2, j \geq m\}.$

Por lo tanto, la sucesión $(\text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{x_{ij} : i \geq n, j \geq m\})_{n\in\mathbb{N}}$ es no decreciente, así que el límite $\text{lím}_{n\to\infty} \text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{x_{ij} : i \geq n, j \geq m\}$ existe.

Proposición 7.14. *Una sucesión doble $(x_{nm})_{n,m\in\mathbb{N}}$ converge a 0 si y sólo si:*

$\text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \sup\{|x_{ij}| : i \geq n, j \geq m\} = 0.$

Demostración

Supongamos que la sucesión doble $(x_{nm})_{n,m\in\mathbb{N}}$ converge a 0. Entonces, dada $\varepsilon > 0$, sea $N \in \mathbb{N}$ tal que $|x_{nm}| < \varepsilon$ para toda pareja de números naturales n y m mayores o iguales a N. Entonces, fijando $n \geq N$, se tiene $\text{lím}_{m\rightsquigarrow\infty} \sup\{|x_{ij}| : i \geq n, j \geq m\} \leq \varepsilon$; por lo tanto $\text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \sup\{|x_{ij}| : i \geq n, j \geq m\} \leq \varepsilon$. Como esto se cumple cualquiera que sea $\varepsilon > 0$, se tiene $\text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \sup\{|x_{ij}| : i \geq n, j \geq m\} = 0$.

Inversamente, supongamos que $\text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightarrow\infty} \sup\{|x_{ij}| : i \geq n, j \geq m\} = 0$. Entonces, dada $\varepsilon > 0$, existe $N_1 \in \mathbb{N}$ tal que $\text{lím}_{m\rightsquigarrow\infty} \sup\{|x_{ij}| : i \geq n, j \geq m\} < \varepsilon$ para cualquier $n \geq N_1$; así que, existe N_2 tal que $\sup\{|x_{ij}| : i \geq N_1, j \geq m\} < \varepsilon$ para cualquier $m \geq N_2$. Por lo tanto, $\sup\{|x_{ij}| : i \geq N_1, j \geq N_2\} < \varepsilon$ y, entonces, $|x_{ij}| < \varepsilon$ para cualquier i y j mayores o iguales que $N = \text{máx}\{N_1, N_2\}$. Así que la sucesión doble $(x_{nm})_{n,m\in\mathbb{N}}$ converge a 0.

■

Proposición 7.15 (**Lema de Fatou para sucesiones dobles**). *Sea* $\{f_{nm}\}_{n,m\in N}$ *una sucesión doble de funciones medibles no negativas. Entonces:*

$$\int_{\mathbb{F}} \text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{f_{ij} : i \geq n, j \geq m\}\, d\mu$$

$$\leq \text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \text{ínf}\left\{\int_{\mathbb{F}} f_{ij} d\mu : i \geq n, j \geq m\right\}.$$

Demostración

Fijando $n \in \mathbb{N}$, la sucesión $(\text{ínf}\{f_{ij} : i \geq n, j \geq m\})_{m\in\mathbb{N}}$ es no decreciente y la sucesión $(\text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{x_{ij} : i \geq n, j \geq m\})_{n\in\mathbb{N}}$ también es no decreciente, así que, por el teorema de la convergencia monótona:

$$\int_{\mathbb{F}} \text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{f_{ij} : i \geq n, j \geq m\}\, d\mu$$

$$= \text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} \text{lím}_{m\rightarrow\infty} \text{ínf}\{f_{ij} : i \geq n, j \geq m\}\, d\mu$$

$$= \text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \int_{\mathbb{F}} \text{ínf}\{f_{ij} : i \geq n, j \geq m\}\, d\mu.$$

Por otra parte, $\text{ínf}\{f_{ij} : i \geq n, j \geq m\} \leq f_{rs}$ para cualesquiera $r \geq n$ y $s \geq m$, así que:

$$\int_{\mathbb{F}} \text{ínf}\{f_{ij} : i \geq n, j \geq m\}\, d\mu \leq \int_{\mathbb{F}} f_{rs} d\mu,$$

para cualesquiera $r \geq n$ y $s \geq m$, por lo tanto:

$$\int_{\mathbb{F}} \text{ínf}\{f_{ij} : i \geq n, j \geq m\}\, d\mu \leq \text{ínf}\left\{\int_{\mathbb{F}} f_{ij} d\mu : i \geq n, j \geq m\right\}.$$

■

Teorema 7.13. *Si* f *es una función medible no negativa tal que* $\int_{\mathbb{F}} f d\mu < \infty$*, entonces* f *es finita casi en todas partes.*

Demostración

Sea $\Gamma = \{x \in \mathbb{F} : f(x) = \infty\}$, entonces $\int_{\mathbb{F}} f I_\Gamma d\mu \leq \int_{\mathbb{F}} f d\mu < \infty$.

Supongamos $\mu(\Gamma) > 0$ y definamos, para cada $n \in \mathbb{N}$,

$$\varphi_n(x) = \begin{cases} n & \text{si } x \in \Gamma \\ 0 & \text{si } x \notin \Gamma \end{cases}$$

Para cada $n \in \mathbb{N}$, φ_n es una función simple tal que $0 \leq \varphi_n \leq f I_\Gamma$; así que $\int_{\mathbb{F}} f I_\Gamma d\mu \geq \int_{\mathbb{F}} \varphi_n d\mu = n\mu(\Gamma)$ para cualquier $n \in \mathbb{N}$. Por lo tanto $\int_{\mathbb{F}} f I_\Gamma d\mu = \infty$, lo cual es una contradicción.

■

7.7. Funciones integrables

Definición 7.7. *Se dice que una función medible f es integrable sobre un conjunto $E \in \Im$ si $\int_E |f| \, d\mu < \infty$.*

Proposición 7.16. *Si f es una función integrable sobre $\mathbb{F}$, entonces f es finita casi en todas partes.*

Demostración
El resultado es un corolario de la proposición 7.13

■

Proposición 7.17. *Una función medible f es integrable sobre un conjunto medible E si y sólo si f^+ y f^- son integrables sobre E.*

Demostración

Se tiene $f^+ \leq |f|$ y $f^- \leq |f|$, así que si f es una función medible integrable sobre E, entonces f^+ y f^- son también integrables sobre E.

Inversamente, si f^+ y f^- son integrables sobre E, entonces $\int_E |f| \, d\mu = \int_E f^+ d\mu + \int_E f^- d\mu$, así que f es también integrable sobre E.

■

Definición 7.8. *Si f una función medible e integrable sobre un conjunto medible E, se define su integral sobre E, $\int_E f d\mu$, de la siguiente manera:*

$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$.

Proposición 7.18. *Sean f y g dos funciones medibles e integrables sobre un conjunto medible E, entonces:*

(i) *Para cualquier número real c, la función cf es integrable sobre E y* $\int_E cf d\mu = c\int_E f d\mu$.
(ii) *Si $f \leq g$ sobre E, entonces* $\int_E f d\mu \leq \int_E g d\mu$.
(iii) $\left|\int_E f d\mu\right| \leq \int_E |f|\, d\mu$

Demostración

1. $|cf| \leq |c|\,|f|$, así que $\int_E |cf|\, d\mu < \infty$.

Si $c < 0$, se tiene:

$(cf)^+ = |c|\, f^-$ y $(cf)^- = |c|\, f^+$, así que:

$\int_E cf d\mu = \int_E |c|\, f^- d\mu - \int_E |c|\, f^+ d\mu$

$= -c\int_E f^- d\mu + c\int_E f^+ d\mu = c\left(\int_E f^- d\mu + \int_E f^+ d\mu\right)$

$= c\int_E f d\mu.$

Si $c \geq 0$, se tiene:

$(cf)^+ = cf^+$y $(af)^- = cf^-$, así que:

$\int_E cf d\mu = \int_E cf^+ d\mu - \int_E cf^- d\mu = c\left(\int_E f^+ d\mu - \int_E f^- d\mu\right)$

$= c\int_E f d\mu.$

2. Si $f \leq g$ sobre E, entonces $g - f \geq 0$ sobre E, así que:

$\int_E g d\mu - \int_E f d\mu = \int_E (g - f)\, d\mu \geq 0.$

Por lo tanto:

$\int_E f d\mu \leq \int_E g d\mu.$

4. $\left|\int_E f d\mu\right| = \left|\int_E f^+ d\mu - \int_E f^- d\mu\right| \leq \int_E f^+ d\mu + \int_E f^- d\mu = \int_E |f|\, d\mu.$

■

PROPOSICIÓN 7.19. *Sean f y g dos funciones medibles e integrables sobre un conjunto medible E y $h : \mathbb{F} \to \overline{\mathbb{R}}$ una función medible tal que $h(x) = f(x) + g(x)$ en todos los puntos $x \in E$ para los cuales $f(x) + g(x)$ esté definida, entonces h es integrable sobre E y:*

$\int_E h d\mu = \int_E f d\mu + \int_E g d\mu.$

Demostración

Sea $\Gamma = \{x \in E : f(x) + g(x) \text{ está definida}\}$.

Como f y g son integrables sobre E, $\mu(E - \Gamma) = 0$, así que:

$\int_E |f| I_{E-\Gamma} d\mu = \int_E |g| I_{E-\Gamma} d\mu = \int_E |h| I_{E-\Gamma} d\mu = 0.$

Por otra parte:

$|hI_\Gamma| = |fI_\Gamma + gI_\Gamma| \leq |fI_\Gamma| + |gI_\Gamma|$, así que $\int_E |h| I_\Gamma d\mu < \infty$.

Además:

$(fI_\Gamma + gI_\Gamma)^+ - (fI_\Gamma + gI_\Gamma)^- = fI_\Gamma + gI_\Gamma = (fI_\Gamma)^+ - (fI_\Gamma)^- + (gI_\Gamma)^+ - (gI_\Gamma)^-$

$= (fI_\Gamma)^+ + (gI_\Gamma)^+ - (fI_\Gamma)^- + (gI_\Gamma)^-.$

Así que:

$(fI_\Gamma + gI_\Gamma)^+ + (fI_\Gamma)^- + (gI_\Gamma)^- = (fI_\Gamma + gI_\Gamma)^- + (fI_\Gamma)^+ + (gI_\Gamma)^+.$

Por lo tanto:

$\int_E (fI_\Gamma + gI_\Gamma)^+ d\mu + \int_E (fI_\Gamma)^- d\mu + \int_E (gI_\Gamma)^- d\mu$

$= \int_E (fI_\Gamma + gI_\Gamma)^- d\mu + \int_E (fI_\Gamma)^+ d\mu + \int_{\mathbb{E}} (gI_\Gamma)^+ d\mu.$

De lo cual se sigue:

$\int_E hI_\Gamma d\mu = \int_E (fI_\Gamma + gI_\Gamma) d\mu = \int_E (fI_\Gamma + gI_\Gamma)^+ d\mu - \int_E (fI_\Gamma + gI_\Gamma)^- d\mu$

$= \int_E (fI_\Gamma)^+ d\mu + \int_E (gI_\Gamma)^+ d\mu - \int_E (fI_\Gamma)^- d\mu - \int_E (gI_\Gamma)^- d\mu$

$= \left(\int_E (fI_\Gamma)^+ d\mu - \int_E (fI_\Gamma)^- d\mu\right) + \left(\int_E (gI_\Gamma)^+ d\mu - \int_E (gI_\Gamma)^- d\mu\right)$

$= \int_E fI_\Gamma d\mu + \int_E gI_\Gamma d\mu.$

■

Un razonamiento de inducción permite demostrar el siguiente corolario:

Corolario 7.7. *Sean $f_1, \ldots, f_n$ n funciones medibles e integrables sobre un conjunto medible E, $a_1, \ldots, a_n$ números reales y $h : \mathbb{F} \to \overline{\mathbb{R}}$ una función medible tal que $h(x) = \sum_{k=1}^n a_k f_k(x)$ en todos los puntos $x \in E$ para los cuales $\sum_{k=1}^n a_k f_k(x)$ esté definida, entonces h es integrable sobre E y:*

$\int_E h d\mu = \sum_{k=1}^n a_k \int_E f_k d\mu..$

Proposición 7.20. *Sean f y g dos funciones medibles e integrables, entonces:*

(i) *Si $\int_E g d\mu \geq 0$ para cualquier $E \in \Im$, entonces $\mu\{x \in \mathbb{F} : g(x) < 0\} = 0$.*
(ii) *$\int_E f d\mu \leq \int_E g d\mu$ para cualquier $E \in \Im$, entonces $\mu\{x \in \mathbb{F} : f(x) > g(x)\} = 0$.*
(iii) *$\int_E f d\mu = \int_E g d\mu$ para cualquier $E \in \Im$, entonces $\mu\{x \in \mathbb{F} : f(x) \neq g(x)\} = 0$.*

Demostración

Supongamos que $\int_E g d\mu \geq 0$ para cualquier $E \in \Im$ y, para cada $n \in \mathbb{N}$, sea $E_n = \left\{x \in \mathbb{F} : g(x) < -\frac{1}{n}\right\}$, entonces, para cualquier $n \in \mathbb{N}$, se tiene:

$0 \leq \int_{E_n} g d\mu \leq -\frac{1}{n}\mu(E_n)$.

Así que $\mu(E_n) = 0$.

Finalmente, $\mu\{x \in \mathbb{F} : g(x) < 0\} = \mu(\cup_{n=1}^{\infty} E_n) = \text{lím}_{n \rightsquigarrow \infty} \mu(E_n) = 0$.

Las otras dos afirmaciones se siguen como corolario de la primera.

■

Pasemos ahora a demostrar el segundo de los teoremas de convergencia de la integral.

TEOREMA 7.14. *Sea g una función no negativa, integrable sobre un conjunto medible E y $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tales que $|f_n| \leq g$ y $f = \text{lím}_{n\rightsquigarrow\infty} f_n$ existe excepto a lo más en un conjunto de medida cero, entonces:*

$\text{lím}_{n\to\infty} \int_E |f_n - f| \, d\mu = 0$.

Demostración

Para cada $n \in \mathbb{N}$, sea $h_n = 2g - |f_n - f|$, entonces, por el lema de Fatou, se tiene:

$2\int_E g d\mu = \int_E \text{lím}_{n\rightsquigarrow\infty} h_n d\mu$

$\leq \text{lím inf}_{n\rightsquigarrow\infty} \int_E h_n d\mu = 2\int_E g d\mu - \text{lím sup}_{n\rightsquigarrow\infty} \int_E |f_n - f| \, d\mu$.

Así que:

$\text{lím sup}_{n\rightsquigarrow\infty} \int_E |f_n - f| \, d\mu = 0$.

■

COROLARIO 7.8 (**Teorema de la convergencia dominada**). *Sea g una función no negativa, integrable sobre un conjunto medible E, y $\{f_n\}_{n\in N}$ una sucesión de funciones medibles tales que $|f_n| \leq g$ y $\text{lím}_{n\rightsquigarrow\infty} f_n$ existe excepto a lo más en un conjunto de medida cero, entonces:*

$\int_E \text{lím}_{n\rightsquigarrow\infty} f_n d\mu = \text{lím}_{n\to\infty} \int_E f_n d\mu$.

Demostración

Sea $f = \text{lím}_{n\rightsquigarrow\infty} f_n$. Entonces:

$\left|\int_E f d\mu - \text{lím}_{n\rightsquigarrow\infty} \int_E f_n d\mu\right| = \text{lím}_{n\rightsquigarrow\infty} \left|\int_E (f_n - f) \, d\mu\right| \leq \text{lím}_{n\to\infty} \int_E |f_n - f| \, d\mu = 0$.

■

Proposición 7.21. *Sea g una función no negativa, integrable y $\{f_{nm}\}_{n,m\in\mathbb{N}}$ una sucesión doble de funciones medibles tales que $|f_{nm}| \leq g$ y que converge a 0 excepto a lo más en un conjunto de medida cero, entonces la sucesión doble $\left\{\int_{\mathbb{F}} |f_{nm}|\, d\mu\right\}_{n,m\in\mathbb{N}}$ converge a 0.*

Demostración

Como $\{f_{nm}\}_{n,m\in\mathbb{N}}$ converge a 0 excepto a lo más en un conjunto de medida cero, entonces:

$\text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \sup\{|f_{ij}| : i \geq n, j \geq m\} = 0,$

excepto a lo más en un conjunto de medida cero. Por lo tanto, si definimos, para cada pareja $n, m \in \mathbb{N}$, $h_{nm} = 2g - |f_{nm}|$, se tiene:

$\text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{h_{ij} : i \geq n, j \geq m\}$

$= \text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{2g - |f_{ij}| : i \geq n, j \geq m\}$

$= 2g - \text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \sup\{|f_{ij}| : i \geq n, j \geq m\} = 2g.$

Entonces, por el lema de Fatou, se tiene:

$2\int_E g d\mu = \int_{\mathbb{F}} \text{lím}_{n\to\infty} \text{lím}_{m\rightsquigarrow\infty} \text{ínf}\{h_{ij} : i \geq n, j \geq m\}\, d\mu$

$\leq \text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \text{ínf}\left\{\int_{\mathbb{F}} h_{ij} d\mu : i \geq n, j \geq m\right\},$

$\text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \text{ínf}\left\{\int_{\mathbb{F}} (2g - |f_{nm}|)\, d\mu : i \geq n, j \geq m\right\}$

$= 2\int_E g d\mu - \text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \sup\left\{\int_{\mathbb{F}} |f_{nm}|\, d\mu : i \geq n, j \geq m\right\}.$

Así que:

$\text{lím}_{n\rightsquigarrow\infty} \text{lím}_{m\rightsquigarrow\infty} \sup\left\{\int_{\mathbb{F}} |f_{nm}|\, d\mu : i \geq n, j \geq m\right\} = 0.$

Por lo tanto, la sucesión doble $\left\{\int_{\mathbb{F}} |f_{nm}|\, d\mu\right\}_{n,m\in\mathbb{N}}$ converge a 0.

■

Teorema 7.15. *Sea f una función medible no negativa, $m : \Im \to \mathbb{R}$ definida por $m(E) = \int_E f d\mu$ y h una función medible e integrable con respecto a m, entonces hf es integrable con respecto a μ y se tiene:*

$\int_{\mathbb{F}} h dm = \int_{\mathbb{F}} hf d\mu.$

Demostración

Se tiene:

$\int_{\mathbb{F}} h^+ dm = \int_{\mathbb{F}} h^+ f d\mu,$

$\int_{\mathbb{F}} h^- dm = \int_{\mathbb{F}} h^- f d\mu.$

Así que h^+f y h^-f son integrables con respecto a μ. Por lo tanto $|hf| = h^+f + h^-f$ es integrable con respecto a μ.

Además:

$\int_{\mathbb{F}} hdm = \int_{\mathbb{F}} h^+dm - \int_{\mathbb{F}} h^-dm = \int_{\mathbb{F}} h^+fd\mu - \int_{\mathbb{F}} h^-fd\mu = \int_{\mathbb{F}} hfd\mu.$

$\blacksquare$

CAPÍTULO 8

TEORÍA GENERAL DE INTEGRACIÓN
Segunda parte

8.1. Integrabilidad uniforme

En esta sección asumiremos que la medida μ es finita.

PROPOSICIÓN 8.1. *Si f es una función medible no negativa, entonces:*

$\int_{\mathbb{F}} f d\mu = \int_0^\infty \mu\left(\{y \in \mathbb{F} : f(y) > x\}\right) dx.$

Demostración

Consideremos primero una función simple no negativa φ con representación canónica $\varphi = \sum_{j=1}^m b_j I_{E_j}$.

En este caso, se tiene:

$\int_0^\infty \mu\left(\{y \in \mathbb{F} : \varphi(y) > x\}\right) dx = \int_0^\infty \sum_{\{j \in \{1,\ldots,m\} : b_j > x\}} \mu(E_j)\, dx$

$= \int_0^\infty \sum_{j=1}^m I_{[0,b_j)}(x) \mu(E_j)\, dx = \sum_{j=1}^m \int_0^\infty I_{[0,b_j)}(x) \mu(E_j)\, dx$

$= \sum_{j=1}^m b_j \mu(E_j) = \int_{\mathbb{F}} \varphi d\mu.$

Consideremos ahora una sucesión no decreciente $(\varphi_n)_{n\in\mathbb{N}}$ de funciones simples no negativas tales que $\lim_{n\rightsquigarrow\infty} \varphi_n(x) = f(x)$ para cualquier $x \in \mathbb{F}$ y, para $n \in \mathbb{N}$ y $x \in [0, \infty)$, definamos $g_n(x) = \mu\left(\{y \in \mathbb{F} : \varphi_n(y) > x\}\right)$.

La sucesión $(g_n)_{n\in\mathbb{N}}$ es no decreciente y $\lim_{n\rightsquigarrow\infty} g_n(x) = \mu\left(\{y \in \mathbb{F} : f(y) > x\}\right)$ para cualquier $x \in [0, \infty)$, así que, por el teorema de la convergencia monótona, se tiene:

$\int_0^\infty \mu\left(\{y \in \mathbb{F} : f(y) > x\}\right) dx = \int_0^\infty \lim_{n\rightsquigarrow\infty} \mu\left(\{y \in \mathbb{F} : \varphi_n(y) > x\}\right) dx$

$= \lim_{n\rightsquigarrow\infty} \int_0^\infty \mu\left(\{y \in \mathbb{F} : \varphi_n(y) > x\}\right) dx = \lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} \varphi_n d\mu = \int_{\mathbb{F}} f d\mu.$

■

TEOREMA 8.1. *Si f es una función integrable, entonces la serie:*

$\sum_{k=1}^{\infty} \mu\left(\{y \in \mathbb{F} : |f(y)| \geq k\}\right)$

converge.

Demostración

$\int_0^{\infty} \mu\left(\{y \in \mathbb{F} : |f(y)| > x\}\right) dx = \lim_{n \rightsquigarrow \infty} \int_0^{\infty} I_{[0,n)}(x)\, \mu\left(\{y \in \mathbb{F} : |f(y)| > x\}\right) dx$

$= \lim_{n \rightsquigarrow \infty} \int_0^{\infty} \sum_{k=1}^{n} I_{[k-1,k)}(x)\, \mu\left(\{y \in \mathbb{F} : |f(y)| > x\}\right) dx$

$= \lim_{n \rightsquigarrow \infty} \sum_{k=1}^{n} \int_0^{\infty} I_{[k-1,k)}(x)\, \mu\left(\{y \in \mathbb{F} : |f(y)| > x\}\right) dx.$

$I_{[k-1,k)}(x)\, \mu\left(\{y \in \mathbb{F} : |f(y)| \geq k\}\right) \leq I_{[k-1,k)}(x)\, \mu\left(\{y \in \mathbb{F} : |f(y)| > x\}\right).$

Así que:

$\sum_{k=1}^{n} \mu\left(\{y \in \mathbb{F} : |f(y)| \geq k\}\right) = \sum_{k=1}^{n} \int_0^{\infty} I_{[k-1,k)}(x)\, \mu\left(\{y \in \mathbb{F} : |f(y)| \geq k\}\right) dx$

$\leq \sum_{k=1}^{n} \int_0^{\infty} I_{[k-1,k)}(x)\, \mu\left(\{y \in \mathbb{F} : |f(y)| > x\}\right) dx.$

Por lo tanto, tomando límites, se obtiene:

$\sum_{k=1}^{\infty} \mu\left(\{y \in \mathbb{F} : |f(y)| \geq k\}\right) \leq \int_0^{\infty} \mu\left(\{y \in \mathbb{F} : |f(y)| > x\}\right) dx = \int_{\mathbb{F}} |f|\, d\mu.$

■

COROLARIO 8.1. *Si f es una función integrable, entonces*

$\lim_{\alpha \rightsquigarrow \infty} \mu\left[|f| > \alpha\right] = 0.$

Demostración

$\lim_{\alpha \rightsquigarrow \infty} \mu\left[|f| > \alpha\right] = \lim_{k \rightsquigarrow \infty} \mu\left(\{y \in \mathbb{F} : |f(y)| \geq k\}\right) = 0.$

■

TEOREMA 8.2. *Si μ es finita, una función medible f es integrable si y sólo si:*

$\lim_{\alpha \rightsquigarrow \infty} \int_{[|f| > \alpha]} |f|\, d\mu = 0.$

Demostración

Supongamos primero que f es integrable. Para cada $n \in \mathbb{N}$, definamos:

$$f_n = \begin{cases} |f| & \text{si } |f| \leq n \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene $\lim_{n \rightsquigarrow \infty} f_n = f$ c.s y $|f_n| \leq |f|$ para cualquier $n \in \mathbb{N}$, así que, por el teorema de la convergencia dominada, se tiene:

$\int_{\mathbb{F}} |f| \, d\mu = \text{lím}_{n\to\infty} \int_{\mathbb{F}} f_n d\mu = \text{lím}_{n\rightsquigarrow\infty} \int_{[|f|\leq n]} |f| \, d\mu$

Por lo tanto:

$\text{lím}_{n\rightsquigarrow\infty} \int_{[|f|>n]} |f| \, d\mu = \text{lím}_{n\rightsquigarrow\infty} \left(\int_{\mathbb{F}} |f| \, d\mu - \int_{[|f|\leq n]} |f| \, d\mu \right) = 0.$

Dada $\varepsilon > 0$, sea $N \in \mathbb{N}$ tal que $\int_{[|f|>n]} |f| \, d\mu < \varepsilon$ para cualquier $n \geq N$, entonces, si $\alpha \geq N$, se tiene:

$\int_{[|f|>\alpha]} |f| \, d\mu \leq \int_{[|f|>N]} |f| \, d\mu < \varepsilon.$

Así que, $\text{lím}_{\alpha\rightsquigarrow\infty} \int_{[|f|>\alpha]} |f| \, d\mu = 0.$

Supongamos ahora que $\text{lím}_{\alpha\rightsquigarrow\infty} \int_{[|f|>\alpha]} |f| \, d\mu = 0$. Entonces, tomando $\alpha > 0$ tal que $\int_{[|f|>\alpha]} |f| \, d\mu < 1$, se tiene:

$\int_{\mathbb{F}} |f| \, d\mu = \int_{[|f|\leq\alpha]} |f| \, d\mu + \int_{[|f|>\alpha]} |f| \, d\mu \leq \alpha\mu\left([|f| \leq \alpha]\right) + 1 < \infty.$

■

TEOREMA 8.3. *Una función medible f es integrable si y sólo si dada $\varepsilon > 0$ existe $\delta > 0$ tal que $\int_A |f| \, d\mu < \varepsilon$ para cualquier conjunto $A \in \Im$ tal que $\mu(A) < \delta$.*

Demostración

Dada $\varepsilon > 0$, tomemos $\alpha > 0$ tal que $\int_{[|f|>\alpha]} |f| \, d\mu < \frac{\varepsilon}{2}$ y $\delta = \frac{\varepsilon}{2\alpha}$. Entonces, para cualquier conjunto $A \in \Im$ tal que $\mu(A) < \delta$, se tiene:

$\int_A |f| \, d\mu = \int_{\{|f|\leq\alpha\}} I_A |f| \, d\mu + \int_{\{|f|>\alpha\}} I_A |f| \, d\mu$

$\leq \alpha \int_{\{|f|\leq\alpha\}} I_A d\mu + \int_{\{|f|>\alpha\}} |f| \, d\mu$

$\leq \alpha \int_{\mathbb{F}} I_A d\mu + \int_{\{|f|>\alpha\}} |f| \, d\mu$

$= \alpha\mu(A) + \int_{\{|f|>\alpha\}} |f| \, d\mu < \varepsilon.$

■

DEFINICIÓN 8.1 (**Integrabilidad uniforme**). *Se dice que una familia $\mathcal{H}$ de funciones medibles es uniformemente integrable si:*

$\text{lím}_{\alpha\rightsquigarrow\infty} \sup \left\{ \int_{[|f|>\alpha]} |f| \, d\mu : f \in \mathcal{H} \right\} = 0.$

TEOREMA 8.4. *Una familia $\mathcal{H}$ de funciones medibles es uniformemente integrable si y sólo si el conjunto $\left\{ \int_{\mathbb{F}} |f| \, d\mu : f \in \mathcal{H} \right\}$ está acotado y, dada cualquier $\varepsilon > 0$, existe $\delta > 0$ tal que $\int_A |f| \, d\mu < \varepsilon$ para cualesquiera $f \in \mathcal{H}$ y $A \in \Im$ tal que $\mu(A) \leq \delta$.*

Demostración

Supongamos primero que la familia $\mathcal{H}$ es uniformemente integrable.

Sea $\alpha > 0$ tal que $\int_{[|f|>\alpha]} |f| \, d\mu < 1$ para cualquier $f \in \mathcal{H}$, se tiene entonces:

$\int_{\mathbb{F}} |f| \, d\mu = \int_{[|f|\leq\alpha]} |f| \, d\mu + \int_{[|f|>\alpha]} |f| \, d\mu \leq \alpha\mu(\mathbb{F}) + 1.$

Así que el conjunto $\left\{\int_{\mathbb{F}} |f| \, d\mu : f \in \mathcal{H}\right\}$ está acotado.

Dada $\varepsilon > 0$, sea $\alpha > 0$ tal que $\int_{[|f|>\alpha]} |f| \, d\mu < \frac{\varepsilon}{2}$ para cualquier $f \in \mathcal{H}$, definamos $\delta = \frac{\varepsilon}{2\alpha}$ y consideremos un conjunto $A \in \Im$ tal que $\mu(A) \leq \delta$. Se tiene entonces:

$\int_A |f| \, d\mu \leq \int_{[|f|\leq\alpha]} I_A |f| \, d\mu + \int_{[|f|>\alpha]} |f| \, d\mu \leq \alpha\mu(A) + \frac{\varepsilon}{2} < \varepsilon.$

Inversamente, supongamos que el conjunto $\left\{\int_{\mathbb{F}} |f| \, d\mu : f \in \mathcal{H}\right\}$ está acotado y, dada cualquier $\varepsilon > 0$, existe $\delta > 0$ tal que $\int_A |f| \, d\mu < \varepsilon$ para cualesquiera $f \in \mathcal{H}$ y $A \in \Im$ tal que $\mu(A) \leq \delta$.

Dada $\varepsilon > 0$, sea $\delta > 0$ tal que $\int_A |f| \, d\mu < \frac{\varepsilon}{2}$ para cualesquiera $f \in \mathcal{H}$ y $A \in \Im$ tal que $\mu(A) \leq \delta$. Definamos $\alpha_0 = \frac{1}{\delta} \sup\left\{\int_{\mathbb{F}} |f| \, d\mu : f \in \mathcal{H}\right\}$ y tomemos $f \in \mathcal{H}$ y $\alpha \geq \alpha_0$. Se tiene entonces:

$\mu([|f| > \alpha]) \leq \frac{1}{\alpha} \int_{\mathbb{F}} |f| \, d\mu \leq \delta.$

Por lo tanto:

$\int_{[|f|>\alpha]} |f| \, d\mu < \frac{\varepsilon}{2}.$

Así que:

$\sup\left\{\int_{[|f|>\alpha]} |f| \, d\mu : f \in \mathcal{H}\right\} < \varepsilon.$

Es decir:

$\lim_{\alpha\rightsquigarrow\infty} \sup\left\{\int_{[|f|>\alpha]} |f| \, d\mu : f \in \mathcal{H}\right\} = 0.$

■

PROPOSICIÓN 8.2. *Sea f una función medible e integrable, Γ un conjunto cualquiera y $\{f_\gamma : \gamma \in \Gamma\}$ una familia de funciones medibles tales que $|f_\gamma| \leq f$ para cualquier $\gamma \in \Gamma$, entonces la familia $\{f_\gamma : \gamma \in \Gamma\}$ es uniformemente integrable.*

Demostración

Para cualquier $\gamma \in \Gamma$, se tiene:

$\int_{[|f_\gamma|>\alpha]} |f_\gamma| \, d\mu \leq \int_{[|f_\gamma|>\alpha]} |f| \, d\mu \leq \int_{[|f|>\alpha]} |f| \, d\mu.$

Así que:

$\lim_{\alpha\rightsquigarrow\infty} \sup\left\{\int_{[|f_\gamma|>\alpha]} |f_\gamma| \, d\mu : \gamma \in \Gamma\right\} \leq \lim_{\alpha\rightsquigarrow\infty} \int_{[|f|>\alpha]} |f| \, d\mu = 0.$

■

Ahora viene el tercer teorema de convergencia de la integral. Es una generalización del teorema de la convergencia dominada ya que, de acuerdo con la proposición anterior, éste es un caso particular del siguiente resultado.

TEOREMA 8.5. *Sea $\{f_n\}_{n\in N}$ una familia uniformemente integrable de funciones medibles tal que $f = \lim_{n\to\infty} f_n$ existe excepto a lo más en un conjunto de medida cero, entonces f es integrable y:*

$\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f| \, d\mu = 0.$

Demostración

Sea $\alpha > 0$ tal que $\int_{[|f_n|>\alpha]} |f_n| \, d\mu < 1$ para cualquier $n \in \mathbb{N}$. Entonces:

$\int_{\mathbb{F}} |f_n| \, d\mu = \int_{[|f|\leq\alpha]} |f_n| \, d\mu + \int_{[|f|>\alpha]} |f_n| \, d\mu$

$\leq \alpha\mu\left([|f_n| \leq \alpha]\right) + 1 \leq \alpha\mu\left(\mathbb{F}\right) + 1.$

Por lo tanto, por el lema de Fatou, se tiene:

$\int_{\mathbb{F}} |f| \, d\mu = \int_{\mathbb{F}} \lim |f_n| \, d\mu \leq \liminf_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n| \, d\mu$

$\leq \alpha\mu\left(\mathbb{F}\right) + 1 < \infty.$

Así que f es integrable.

Para cada $\alpha > 0$, definamos:

$$f_n^{(\alpha)} = \begin{cases} f_n & \text{si } |f_n| \leq \alpha \\ 0 & \text{en otro caso} \end{cases}$$

$$f^{(\alpha)} = \begin{cases} f & \text{si } |f| \leq \alpha \\ 0 & \text{en otro caso} \end{cases}$$

Sea $C = \{x \in \mathbb{F} : \lim_{n\to\infty} f_n(x) = f(x)\}$.

Si $x \in C$ es tal que $|f(x)| < \alpha$, entonces $\lim_{n\rightsquigarrow\infty} f_n^{(\alpha)}(x) = f^{(\alpha)}(x)$. Así que, si $\mu\left([|f| = \alpha]\right) = 0$, entonces $\lim_{n\rightsquigarrow\infty} f_n^{(\alpha)} = f^{(\alpha)}$ c.s y, como $\left|f_n^{(\alpha)} - f^{(\alpha)}\right| \leq 2\alpha$, se tiene, por el teorema de la convergencia dominada:

$\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} \left|f_n^{(\alpha)} - f^{(\alpha)}\right| d\mu = 0.$

Por otra parte:

$\int_{\mathbb{F}} \left|f_n^{(\alpha)} - f^{(\alpha)}\right| d\mu$

$= \int_{[|f_n|\leq\alpha,|f|\leq\alpha]} |f_n - f| \, d\mu + \int_{[|f_n|\leq\alpha,|f|>\alpha]} |f_n| \, d\mu + \int_{[|f_n|>\alpha,|f|\leq\alpha]} |f| \, d\mu.$

$\int_{\mathbb{F}} |f_n - f|\, d\mu = \int_{[|f_n|\leq\alpha, |f|\leq\alpha]} |f_n - f|\, d\mu + \int_{[|f_n|\leq\alpha, |f|>\alpha]} |f_n - f|\, d\mu$

$+ \int_{[|f_n|>\alpha, |f|\leq\alpha]} |f_n - f|\, d\mu + \int_{[|f_n|>\alpha, |f|>\alpha]} |f_n - f|\, d\mu$

$\leq \int_{[|f_n|\leq\alpha, |f|\leq\alpha]} |f_n - f|\, d\mu.$

$+ \int_{[|f_n|\leq\alpha, |f|>\alpha]} |f_n|\, d\mu + \int_{[|f_n|\leq\alpha, |f|>\alpha]} |f|\, d\mu$

$+ \int_{[|f_n|>\alpha, |f|\leq\alpha]} |f_n|\, d\mu + \int_{[|f_n|>\alpha, |f|\leq\alpha]} |f|\, d\mu$

$+ \int_{[|f_n|>\alpha, |f|>\alpha]} |f_n|\, d\mu + \int_{[|f_n|>\alpha, |f|>\alpha]} |f|\, d\mu$

$= \int_{\mathbb{F}} \left| f_n^{(\alpha)} - f^{(\alpha)} \right| d\mu$

$+ \int_{[|f_n|\leq\alpha, |f|>\alpha]} |f|\, d\mu + \int_{[|f_n|>\alpha, |f|>\alpha]} |f|\, d\mu$

$+ \int_{[|f_n|>\alpha, |f|\leq\alpha]} |f_n|\, d\mu + \int_{[|f_n|>\alpha, |f|>\alpha]} |f_n|\, d\mu$

$= \int_{\mathbb{F}} \left| f_n^{(\alpha)} - f^{(\alpha)} \right| d\mu + \int_{[|f|>\alpha]} |f|\, d\mu + \int_{[|f_n|>\alpha]} |f_n|\, d\mu.$

Dada $\varepsilon > 0$, sea α_0 tal que $\int_{[|f|>\alpha]} |f|\, d\mu < \varepsilon$ y $\int_{[|f_n|>\alpha]} |f_n|\, d\mu < \varepsilon$, para cualquier $\alpha \geq \alpha_0$, entonces:

$\int_{\mathbb{F}} |f_n - f|\, d\mu \leq \int_{\mathbb{F}} \left| f_n^{(\alpha)} - f^{(\alpha)} \right| d\mu + 2\varepsilon,$

para cualquier $\alpha \geq \alpha_0$.

Sea $\alpha \geq \alpha_0$ tal que $\mu\left([|f| = \alpha]\right) = 0$, entonces:

$\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu \leq 2\varepsilon.$

Así que, como $\varepsilon > 0$ es arbitraria, se tiene:

$\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = 0.$

■

COROLARIO 8.2 (**Teorema de la convergencia uniformemente integrable**). *Sea* $\{f_n\}_{n\in N}$ *una familia uniformemente integrable de funciones medibles tales que* $\lim_{n\to\infty} f_n$ *existe excepto a lo más en un conjunto de medida cero, entonces* $f = \lim_{n\rightsquigarrow\infty} f_n$ *es integrable y:*

$\int_{\mathbb{F}} \lim_{n\rightsquigarrow\infty} f_n d\mu = \lim_{n\to\infty} \int_{\mathbb{F}} f_n d\mu.$

Tenemos un inverso del teorema 8.5:

TEOREMA 8.6. *Sean* $(f_n)_{n\in N}$ *una sucesión de funciones integrables y* f *una función medible tales que* $\lim_{n\to\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = 0$. *Entonces la familia* $\{f_n : n \in \mathbb{N}\}$ *es uniformemente integrable.*

Demostración

Sea $N \in \mathbb{N}$ tal que $\int_{\mathbb{F}} |f_N - f|\, d\mu < 1$, entonces:

$\int_{\mathbb{F}} |f|\, d\mu \leq \int_{\mathbb{F}} |f_N - f|\, d\mu + \int_{\mathbb{F}} |f_N|\, d\mu < \infty$.

Así que f es integrable.

Dada $\varepsilon > 0$, sea $N \in \mathbb{N}$ tal que $\int_{\mathbb{F}} |f_n - f|\, d\mu < \frac{\varepsilon}{2}$ para cualquier $n > N$, se tiene entonces, para cualquier $n \in \mathbb{N}$:

$\int_{\mathbb{F}} |f_n|\, d\mu \leq \int_{\mathbb{F}} |f|\, d\mu + \int_{\mathbb{F}} |f_n - f|\, d\mu$

$< \int_{\mathbb{F}} |f|\, d\mu + \text{máx}\left\{\int_{\mathbb{F}} |f_1 - f|\, d\mu, \ldots, \int_{\mathbb{F}} |f_N - f|\, d\mu, \frac{\varepsilon}{2}\right\}$.

Así que el conjunto $\left\{\int_{\mathbb{F}} |f_n|\, d\mu : n \in \mathbb{N}\right\}$ está acotado.

Tomemos ahora $\delta > 0$ tal que $\text{máx}\left\{\int_A |f|\, d\mu, \int_A |f_1|\, d\mu, \ldots, \int_A |f_N|\, d\mu\right\} < \frac{\varepsilon}{2}$ para cualquier conjunto $A \in \Im$ tal que $\mu(A) < \delta$. Entonces, si $\mu(A) < \delta$ y $n > N$, se tiene:

$\int_A |f_n|\, d\mu \leq \int_A |f|\, d\mu + \int_{\mathbb{F}} |f_n - f|\, d\mu < \varepsilon$.

■

TEOREMA 8.7. *Sea $(f_n)_{n\in N}$ una sucesión de funciones integrables no negativas tal que $f = \text{lím}_{n\rightsquigarrow\infty} f_n$ existe excepto a lo más en un conjunto de medida cero y $\text{lím}_{n\to\infty} \int_{\mathbb{F}} f_n d\mu = \int_{\mathbb{F}} f d\mu < \infty$. Entonces* $\text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = 0$.

Demostración

La sucesión $\{\text{mín}(f_n, f)\}_{n\in N}$ está acotada por f y converge puntualmente a f excepto a lo más en un conjunto de medida cero, así que, por el teorema de la convergencia dominada, se tiene que $\text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} \text{mín}(f_n, f)\, d\mu = \int_{\mathbb{F}} f d\mu$. Por lo tanto:

$\text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} \text{máx}(f_n, f)\, d\mu = \text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} [f_n + f - \text{mín}(f_n, f)]\, d\mu$

$= \text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} f_n d\mu + \int_{\mathbb{F}} f d\mu - \text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} \text{mín}(f_n, f)\, d\mu = \int_{\mathbb{F}} f d\mu$.

Así que:

$\text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = \text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} [\text{máx}(f_n, f) - \text{mín}(f_n, f)]\, d\mu = 0$.

■

COROLARIO 8.3. *Sea $\{f_n\}_{n\in N}$ una sucesión de funciones integrables no negativas tal que $f = \text{lím}_{n\rightsquigarrow\infty} f_n$ existe excepto a lo más en un conjunto de medida cero y $\text{lím}_{n\to\infty} \int_{\mathbb{F}} f_n d\mu = \int_{\mathbb{F}} f d\mu < \infty$. Entonces la familia $\{f_n : n \in \mathbb{N}\}$ es uniformemente integrable.*

Combinando los teoremas 8.5, 8.6, y 8.7 y el corolario 8.2, se tiene el siguiente resultado:

Teorema 8.8. *Sea* $(f_n)_{n\in N}$ *una sucesión de funciones integrables tales que* $f = \lim_{n\rightsquigarrow\infty} f_n$ *existe excepto a lo más en un conjunto de medida cero, entonces las siguientes tres condiciones son equivalentes:*

(i) *La familia* $\{f_n : n \in \mathbb{N}\}$ *es uniformemente integrable.*
(ii) $\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = 0$
(iii) $\lim_{n\to\infty} \int_{\mathbb{F}} |f_n|\, d\mu = \int_{\mathbb{F}} |f|\, d\mu < \infty$

Demostración

Si la familia $\{f_n : n \in \mathbb{N}\}$ es uniformemente integrable, entonces, por el teorema 8.5, se tiene $\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = 0$. Así que la primera condición implica la segunda.

Si $\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = 0$, entonces, por el teorema 8.6, la familia $\{f_n : n \in \mathbb{N}\}$ es uniformemente integrable. Así que la segunda condición implica la primera.

Si $\lim_{n\to\infty} \int_{\mathbb{F}} |f_n|\, d\mu = \int_{\mathbb{F}} |f|\, d\mu < \infty$, entonces, por el teorema 8.7, se tiene:

$\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = 0$.

Así que la tercera condición implica la segunda.

Si a familia $\{f_n : n \in \mathbb{N}\}$ es uniformemente integrable, entonces la familia $\{|f_n| : n \in \mathbb{N}\}$ también es uniformemente integrable, así que, por el corolario 8.2, se tiene:

$\lim_{n\to\infty} \int_{\mathbb{F}} |f_n|\, d\mu = \int_{\mathbb{F}} |f|\, d\mu < \infty$.

Por lo tanto, la primera condición implica la tercera.

■

Corolario 8.4. *Sea* $\{f_n\}_{n\in N}$ *una sucesión de funciones integrables no negativas tales que* $f = \lim_{n\rightsquigarrow\infty} f_n$ *existe excepto a lo más en un conjunto de medida cero, entonces las siguientes tres condiciones son equivalentes:*

(i) *La familia* $\{f_n : n \in \mathbb{N}\}$ *es uniformemente integrable.*
(ii) $\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = 0$
(iii) $\lim_{n\to\infty} \int_{\mathbb{F}} f_n d\mu = \int_{\mathbb{F}} f d\mu < \infty$

Teorema 8.9. *Sea* $\mathcal{H}$ *una familia de funciones medibles uniformemente integrable. Entonces la familia:*

$\mathcal{G} = \Big\{ f : \mathbb{F} \to \overline{\mathbb{R}}^+$: *Existe una sucesión* $(f_n)_{n\in\mathbb{N}}$ *de funciones en* $\mathcal{H}$ *tales que*

$\lim_{n\to\infty} f_n = f$ *excepto a lo más en un conjunto de medida cero*$\}$

es uniformemente integrable.

Demostración

Primero observemos que si $A \in \Im$, entonces la familia de funciones $\mathcal{H}' = \{fI_A : f \in \mathcal{H}\}$ es uniformemente integrable.

Sea M una cota del conjunto $\left\{\int_{\mathbb{F}} |f|\, d\mu : f \in \mathcal{H}\right\}$.

Sea $f \in \mathcal{G}$ y $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones en $\mathcal{H}$ tales que $\lim_{n\to\infty} f_n = f$ excepto a lo más en un conjunto de medida cero, entonces, por la proposición 8.5, f es integrable y: $\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = 0$. Sea $N \in \mathbb{N}$ tal que $\int_{\mathbb{F}} |f_N - f|\, d\mu < 1$. Entonces:

$\int_{\mathbb{F}} |f|\, d\mu \leq \int_{\mathbb{F}} |f_N - f|\, d\mu + \int_{\mathbb{F}} |f_N|\, d\mu < 1 + M.$

Así que la familia $\left\{\int_{\mathbb{F}} |f|\, d\mu : f \in \mathcal{G}\right\}$ está acotada.

Ahora, dada cualquier $\varepsilon > 0$, sea $\delta > 0$ tal que $\int_A |f|\, d\mu < \frac{1}{2}\varepsilon$ para cualesquiera $f \in \mathcal{H}$ y $A \in \Im$ tal que $\mu(A) \leq \delta$.

Si $f \in \mathcal{G}$, sea $A \in \Im$ tal que $\mu(A) \leq \delta$ y $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones en $\mathcal{H}$ tales que $\lim_{n\to\infty} f_n = f$ excepto a lo más en un conjunto de medida cero, entonces $\lim_{n\to\infty} f_n I_A = fI_A$ excepto a lo más en un conjunto de medida cero, así que, por la proposición 8.5:

$\lim_{n\rightsquigarrow\infty} \int_{\mathbb{A}} |f_n - f|\, d\mu = \lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n I_A - fI_A|\, d\mu = 0.$

Sea $N \in \mathbb{N}$ tal que $\int_{\mathbb{A}} |f_N - f|\, d\mu < \frac{1}{2}\varepsilon$.

Entonces:

$\int_{\mathbb{A}} |f|\, d\mu \leq \int_{\mathbb{A}} |f_N - f|\, d\mu + \int_{\mathbb{A}} |f_N|\, d\mu < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon.$

Así que, por la proposición 8.4, $\mathcal{G}$ es uniformemente integrable.

■

8.2. Teorema de Radon-Nikodym

Dadas dos medidas μ y ν, ambas definidas sobre $\Im$, ¿bajo que condiciones existe una función medible no negativa f tal que $\nu(E) = \int_E f d\mu$ para cualquier $E \in \Im$? En esta sección se dará respuesta a esta pregunta.

Dada una función medible $f : \mathbb{F} \to \overline{\mathbb{R}}$, la familia de conjuntos medibles $B_\alpha = [f \leq \alpha]$, con $\alpha \in \mathbb{R}$, es creciente en el sentido de que si $\alpha < \beta$ entonces $B_\alpha \subset B_\beta$. Cabe entonces preguntarse si, dada una familia creciente de conjuntos medibles $\{B_\alpha\}_{\alpha\in\mathbf{R}}$, existe una función medible $f : \mathbb{F} \to \overline{\mathbb{R}}$ tal que $B_\alpha = [f \leq \alpha]$ para cualquier $\alpha \in \mathbb{R}$.

Para definir tal función, dado un punto $x \in \mathbb{F}$, se debe de tener $f(x) \leq \alpha$ para cualquier α tal que $x \in B_\alpha$, así que:

$f(x) \leq \text{ínf}\,\{\alpha \in \mathbb{R} : x \in B_\alpha\}$.

Además, si $f(x) < \text{ínf}\,\{\alpha \in \mathbb{R} : x \in B_\alpha\}$, entonces podría existir β tal que $x \notin B_\beta$ y $f(x) < \beta < \text{ínf}\,\{\alpha \in \mathbb{R} : x \in B_\alpha\}$. En este caso se tendría $f(x) \leq \beta$, pero $x \notin B_\beta$, así que $B_\beta \neq [f \leq \beta]$.

La definición natural de f es entonces $f(x) = \text{ínf}\,\{\alpha \in \mathbb{R} : x \in B_\alpha\}$, aunque con esta definición f podría no ser medible. El siguiente lema precisa el resultado que se tiene en este sentido.

LEMA 8.1. *Sea D un conjunto numerable de números reales y $\{B_\alpha\}_{\alpha \in D}$ una familia de conjuntos medibles tales que si $\alpha < \beta$ entonces $B_\alpha \subset B_\beta$. Entonces existe una función medible $f : \mathbb{F} \to \overline{\mathbb{R}}$ tal que $[f < \alpha] \subset B_\alpha \subset [f \leq \alpha]$ para cualquier $\alpha \in D$.*

Demostración

Para cada $x \in \mathbb{F}$, sea $f(x) = \text{ínf}\,\{\alpha \in D : x \in B_\alpha\}$. Inmediatamente se tiene la contención $B_\alpha \subset [f \leq \alpha]$ para cualquier $\alpha \in D$.

Ahora bien, si α es cualquier número real y $f(x) < \alpha$, entonces existe $\beta < \alpha$ tal que $\beta \in D$ y $x \in B_\beta$, así que $[f < \alpha] \subset \bigcup_{\{\beta \in D : \beta < \alpha\}} B_\beta$. Por otra parte, si $x \in B_\beta$ para alguna $\beta \in D$ con $\beta < \alpha$, entonces $f(x) \leq \beta < \alpha$, así que $\bigcup_{\{\beta \in D : \beta < \alpha\}} B_\beta \subset [f < \alpha]$. Por lo tanto, $[f < \alpha] = \bigcup_{\{\beta \in D : \beta < \alpha\}} B_\beta$, lo cual muestra que f es medible y que $[f < \alpha] \subset B_\alpha$ para cualquier $\alpha \in D$.

■

DEFINICIÓN 8.2. *Sean μ y ν dos medidas. Se dice que ν es absolutamente continua con respecto a μ, lo cual será denotado por $\nu \ll \mu$, si $\nu(E) = 0$ para cualquier conjunto medible E tal que $\mu(E) = 0$.*

Si μ es una medida y f una función medible no negativa, entonces $\nu : \Im \to \mathbb{R}$ definida por $\nu(E) = \int_E f d\mu$ es una medida absolutamente continua con respecto a μ.

TEOREMA 8.10 (**Radon-Nikodym**). *Sean μ y ν dos medidas. Supongamos que μ es finita y que ν es absolutamente continua con respecto a μ, entonces existe una función medible no negativa f tal que $\nu(E) = \int_E f d\mu$ para cualquier conjunto medible E.*

Demostración

Para cada número racional definamos la medida con signo $\nu_\alpha = v - \alpha\mu$ y sea (A_α, B_α) una descomposición de Hahn para ν_α. Para $\alpha = 0$ tomemos $A_0 = X$ y $B_0 = \emptyset$.

Sea $\alpha < \beta$ entonces, como $B_\alpha - B_\beta = B_\alpha \cap A_\beta$, se tiene $B_\alpha - B_\beta \subset A_\beta$ y $B_\alpha - B_\beta \subset B_\alpha$. Por lo tanto, $\nu_\beta(B_\alpha - B_\beta) \geq 0$ y $\nu_\alpha(B_\alpha - B_\beta) \leq 0$, es decir, $v(B_\alpha - B_\beta) - \beta\mu(B_\alpha - B_\beta) \geq 0$ y $v(B_\alpha - B_\beta) - \alpha\mu(B_\alpha - B_\beta) \leq 0$, así que $\beta\mu(B_\alpha - B_\beta) \leq v(B_\alpha - B_\beta) \leq \alpha\mu(B_\alpha - B_\beta)$, de lo cual se sigue $\mu(B_\alpha - B_\beta) = 0$.

Sea $F = \bigcup_{\{\alpha,\beta\in\mathbf{Q}:\alpha<\beta\}}(B_\alpha - B_\beta)$. Entonces $\mu(F) = 0$ y $B_\alpha - F \subset B_\beta - F$ para cualquier pareja de racionales α y β tales que $\alpha < \beta$.

Definamos $B'_\alpha = B_\alpha - F$ y $A'_\alpha = (B'_\alpha)^c$ para cualquier número racional α. Entonces A_α y B_α difieren de A'_α y B'_α, respectivamente, por un conjunto de medida μ igual a cero, el cual también tiene medida ν igual a cero ya que ν es absolutamente continua con respecto a μ. Esto implica que, para cada α racional, la pareja (A'_α y B'_α) es también una descomposición de Hahn para ν_α. Obsérvese, además, que se sigue teniendo $A'_0 = X$ y $B'_0 = \emptyset$.

Para simplificar la notación, se puede asumir que esta nueva descomposición de Hahn es la que se toma inicialmente. De esta forma, la familia $\{B_\alpha\}_{\alpha\in\mathbf{Q}}$ es creciente.

Por el lema 8.1 existe una función medible $f : \mathbb{F} \to \overline{\mathbb{R}}$ tal que, para cualquier número racional α, $[f < \alpha] \subset B_\alpha \subset [f \leq \alpha]$.

Como $B_0 = \emptyset$, $[f < 0] = \emptyset$, así que f es no negativa.

Sean $\alpha, \beta \in \mathbb{Q}$ con $\alpha < \beta$, entonces $\alpha \leq f \leq \beta$ sobre $B_\beta - B_\alpha = B_\beta \cap A_\alpha$, así que, si $E \in \Im$ y $E_{\alpha\beta} = E \cap (B_\beta - B_\alpha)$, entonces:

$$\alpha\mu(E_{\alpha\beta}) \leq \int_{E_{\alpha\beta}} f d\mu \leq \beta\mu(E_{\alpha\beta}). \tag{8.1}$$

Por otra parte, como $E_{\alpha\beta} \subset B_\beta \cap A_\alpha$, se tiene $\nu_\alpha(E_{\alpha\beta}) \geq 0$ y $\nu_\beta(E_{\alpha\beta}) \leq 0$, es decir, $\nu(E_{\alpha\beta}) - \alpha\mu(E_{\alpha\beta}) \geq 0$ y $\nu(E_{\alpha\beta}) - \beta\mu(E_{\alpha\beta}) \leq 0$. Así que:

$$\alpha\mu(E_{\alpha\beta}) \leq \nu(E_{\alpha\beta}) \leq \beta\mu(E_{\alpha\beta}). \tag{8.2}$$

Combinando las desigualdades 8.1 y 8.2, se obtiene:

$$\nu(E_{\alpha\beta}) - (\beta - \alpha)\mu(E_{\alpha\beta}) \leq \int_{E_{\alpha\beta}} f d\mu \leq \nu(E_{\alpha\beta}) + (\beta - \alpha)\mu(E_{\alpha\beta}). \tag{8.3}$$

En particular, si, para cada $N \in \mathbb{N}$, consideramos los conjuntos $B_0, B_{\frac{1}{N}}, B_{\frac{2}{N}}, \ldots$ y, para cada $k \in \{0, 1, 2, \ldots\}$, definimos $E_k = E \cap \left(B_{\frac{k+1}{N}} - B_{\frac{k}{N}}\right)$, aplicando 8.3 para cada k, se tiene:

$$\nu(E_k) - \frac{1}{N}\mu(E_k) \leq \int_{E_k} f d\mu \leq \nu(E_k) + \frac{1}{N}\mu(E_k). \tag{8.4}$$

De manera que, como los conjuntos E_k son ajenos por parejas, sumando sobre k cada término de la desigualdad 8.4, se obtiene:

$$(8.5) \qquad \nu(\bigcup_{k=0}^{\infty} E_k) - \frac{1}{N}\mu(\bigcup_{k=0}^{\infty} E_k) \leq \int_{\bigcup_{k=0}^{\infty} E_k} f d\mu \leq \nu(\bigcup_{k=0}^{\infty} E_k) + \frac{1}{N}\mu(\bigcup_{k=0}^{\infty} E_k).$$

Sea ahora $E_\infty = E - \bigcup_{k=0}^{\infty} B_{\frac{k}{N}}$.

Como $[f < \infty] \subset \bigcup_{k=0}^{\infty} B_{\frac{k}{N}}$, se tiene $f = \infty$ sobre E_∞. Por otra parte, para cualquier $k \in \{0,1,2,\ldots\}$, se tiene $E_\infty \subset A_{\frac{k}{N}}$, así que $\nu(E_\infty) - \frac{k}{N}\mu(E_\infty) \geq 0$, es decir, $\nu(E_\infty) \geq \frac{k}{N}\mu(E_\infty)$. Por lo tanto, si $\mu(E_\infty) > 0$, entonces $\nu(E_\infty) = \infty$. Por otra parte, si $\mu(E_\infty) = 0$, entonces $\nu(E_\infty) = 0$ ya que ν es absolutamente continua con respecto a μ. Por lo tanto, en cualquier caso se tiene:

$$(8.6) \qquad \nu(E_\infty) = \int_{E_\infty} f d\mu.$$

Finalmente, como $E = E_\infty \cup [\bigcup_{k=0}^{\infty} E_k]$ y los conjuntos E_∞ y $\bigcup_{k=0}^{\infty} E_k$ son ajenos, se tiene $\nu(E) = \nu(E_\infty) + \nu(\bigcup_{k=0}^{\infty} E_k)$ y $\int_E f d\mu = \int_{E_\infty} f d\mu + \int_{\bigcup_{k=0}^{\infty} E_k} f d\mu$. Además, $\mu(\bigcup_{k=0}^{\infty} E_k) \leq \mu(E)$. Así que, combinando 8.5 y 8.6, se obtiene:

$$\nu(E) - \frac{1}{N}\mu(E) \leq \int_E f d\mu \leq \nu(E) + \frac{1}{N}\mu(E).$$

Siendo μ finita y $N \in \mathbb{N}$ arbitraria, se concluye $\nu(E) = \int_E f d\mu$.

■

COROLARIO 8.5. *Sean μ y ν dos medidas. Supongamos que μ es σ-finita y que ν es absolutamente continua con respecto a μ, entonces existe una función medible no negativa f tal que $\nu(E) = \int_E f d\mu$ para cualquier conjunto medible E.*

Demostración

Sea $\{E_k\}_{k\in\mathbb{N}}$ una colección infinita numerable de conjuntos, ajenos por parejas, $E_k \in \Im$ tales que $\mathbb{F} = \bigcup_{k=1}^{\infty} E_k$ y $\mu(E_k) < \infty$ para cualquier k.

Para cada $k \in \mathbb{N}$, sean $\mu_k : \Im \mapsto \mathbb{R}$ y $\nu_k : \Im \mapsto \overline{\mathbb{R}}$ definidas por:

$\mu_k(E) = \mu(E \cap E_k)$,

$\nu_k(E) = \nu(E \cap E_k)$.

μ_k y ν_k son medidas y μ_k es finita.

Además, si $E \in \Im$ y $\mu_k(E) = 0$, entonces $\mu(E \cap E_k)$, así que $\nu_k(E) = \nu(E \cap E_k) = 0$.

Por lo tanto, ν_k es absolutamente continua con respecto a μ_k. Así que existe una función medible no negativa f_k tal que $\nu_k(E) = \int_E f_k d\mu_k$ para cualquier conjunto medible E.

Como $\mu_k(E_k^c) = 0$, podemos redefinir f_k de tal forma que sea nula sobre E_k^c, así que $\nu_k(E) = \int_E f_k d\mu$.

Sea $f = \sum_{k=1}^{\infty} f_k$, entonces f es una función medible no negativa y, para cualquier conjunto medible E, se tiene:

$\nu(E) = \sum_{k=1}^{\infty} \nu(E \cap E_k) = \sum_{k=1}^{\infty} \nu_k(E)$

$= \sum_{k=1}^{\infty} \int_E f_k d\mu = \int_E \sum_{k=1}^{\infty} f_k d\mu = \int_E f d\mu.$

■

8.3. Producto de espacios de medida

En esta sección, $(\mathbb{F}_1, \Im_1, \mu_1)$ y $(\mathbb{F}_2, \Im_2, \mu_2)$ serán dos espacios de medida cualesquiera.

DEFINICIÓN 8.3. *Por un rectángulo medible en $\mathbb{F}_1 \times \mathbb{F}_2$ se entenderá un conjunto de la forma $A \times B$, donde $A \in \Im_1$ y $B \in \Im_2$. La σ-álgebra generada por los rectángulos medibles en $\mathbb{F}_1 \times \mathbb{F}_2$ será llamada la σ-álgebra producto de $\Im_1$ y $\Im_2$ y será denotada por $\Im_1 \times \Im_2$.*

Sea $\mathcal{A}$ la familia de conjuntos de la forma $\bigcup_{j=1}^{n} R_j$ en donde $n \in \mathbb{N}$ y $R_1, \ldots, R_n$ son rectángulos medibles en $\mathbb{F}_1 \times \mathbb{F}_2$, ajenos por parejas. Obviamente $\mathcal{A}$ es un álgebra de subconjuntos de $\mathbb{F}_1 \times \mathbb{F}_2$ y la σ-álgebra generada por $\mathcal{A}$ es $\Im_1 \times \Im_2$.

Si $R = A \times B$ es un rectángulo medible en $\mathbb{F}_1 \times \mathbb{F}_2$, definamos:

$\mu_0(R) = \mu_1(A)\,\mu_2(B)$.

Si $E = \bigcup_{j=1}^{m} R^{(j)} \in \mathcal{A}$, definamos:

$\mu_0(E) = \sum_{j=1}^{m} \mu_0\left(R^{(j)}\right)$.

Evidentemente, la función $\mu_0 : \mathcal{A} \mapsto \mathbb{R} \cup \{\infty\}$ es finitamente aditiva y $\mu_0(\emptyset) = 0$.

PROPOSICIÓN 8.3. *Sea $R = A \times B$ un rectángulo medible en $\mathbb{F}_1 \times \mathbb{F}_2$ y $R_i = A_i \times B_i$ una colección infinita numerable de rectángulos medibles en $\mathbb{F}_1 \times \mathbb{F}_2$, ajenos por parejas, tal que $R = \bigcup_{k=1}^{\infty} R_i$, entonces:*

$\mu_0(R) = \sum_{i=1}^{\infty} \mu_0(R_i)$.

Demostración

Para cada $x \in A$, se tiene:

$B = \bigcup_{\{i: x \in A_i\}} B_i$.

Por lo tanto:

$\mu_2(B) = \sum_{\{i:x\in A_i\}} \mu_2(B_i) = \sum_{i=1}^{\infty} \mu_2(B_i) I_{A_i}(x)$.

Así que:

$\mu_2(B) I_A = \sum_{i=1}^{\infty} \mu_2(B_i) I_{A_i}$.

Integrando, se obtiene:

$\mu_2(B)\mu_1(A) = \int_{\mathbb{F}_1} \mu_2(B) I_A d\mu_1 = \int_{\mathbb{F}_1} \sum_{i=1}^{\infty} \mu_2(B_i) I_{A_i} d\mu_1$

$= \sum_{i=1}^{\infty} \int_{\mathbb{F}_1} \mu_2(B_i) I_{A_i} d\mu_1 = \sum_{i=1}^{\infty} \mu_2(B_i)\mu_1(A_i)$.

■

Teorema 8.11. *μ_0 es una quasi medida.*

Demostración

Sea $E_1, E_2, \ldots$ una colección infinita numerable de elementos de $\mathcal{A}$, ajenos por parejas, tal que $E = \bigcup_{i=1}^{\infty} E_i \in \mathcal{A}$.

Por un lado, como $E = \bigcup_{i=1}^{\infty} E_i$, E es una unión infinita numerable de rectángulos medibles R_k en $\mathbb{F}_1 \times \mathbb{F}_2$, ajenos por parejas.

Por otro lado, como $E \in \mathcal{A}$, E es una unión finita de rectángulos medibles en $\mathbb{F}_1 \times \mathbb{F}_2$, ajenos por parejas.

Sea $E = \bigcup_{j=1}^{m} R^{(j)}$.

Para cada $j \in \{1, \ldots, m\}$ y $k \in \mathbb{N}$, definamos $R_k^{(j)} = R_k \cap R^{(j)}$. Entonces, como $\bigcup_{j=1}^{m} R^{(j)} = \bigcup_{k=1}^{\infty} R_k$, se tiene $R_k = \bigcup_{j=1}^{m} R_k^{(j)}$ y $R^{(j)} = \bigcup_{k=1}^{\infty} R_k^{(j)}$, así que:

$\mu_0(E) = \sum_{j=1}^{m} \mu_0\left(R^{(j)}\right) = \sum_{j=1}^{m} \sum_{k=1}^{\infty} \mu_0\left(R_k^{(j)}\right)$

$= \sum_{k=1}^{\infty} \sum_{j=1}^{m} \mu_0\left(R_k^{(j)}\right) = \sum_{k=1}^{\infty} \mu_0(R_k) = \sum_{i=1}^{\infty} \mu_0(E_i)$.

■

De acuerdo con el teorema 5.6, la quasi medida μ_0 puede extenderse a una medida definida sobre la σ-álgebra $(\Im_1 \times \Im_2)^*$ generada por $\mathcal{A}$ y los conjuntos de medida exterior cero definidos por μ_0. Denotaremos a esa medida por $\mu_1 \times \mu_2$ y la llamaremos la medida producto de μ_1 y μ_2.

Si $E \in \Im_1 \times \Im_2$ definamos, para cada $x \in \mathbb{F}_1$:

$E_x = \{y \in \mathbb{F}_2 : (x, y) \in E\}$, y, para cada $y \in \mathbb{F}_2$:

$E^y = \{x \in \mathbb{F}_1 : (x, y) \in E\}$.

PROPOSICIÓN 8.4. *$E_x \in \Im_2$ para cualquier $E \in \Im_1 \times \Im_2$ y $x \in \mathbb{F}_1$.*

Demostración

Sea $\mathcal{H} = \{E \in \Im_1 \times \Im_2 : E_x \in \Im_2$ para cualquier $x \in \mathbb{F}_1\}$.

$\mathcal{H}$ es una σ-álgebra que contiene a los rectángulos medibles en $\mathbb{F}_1 \times \mathbb{F}_2$, así que $\mathcal{H}$ contiene a $\Im_1 \times \Im_2$.

■

COROLARIO 8.6. *$E^y \in \Im_1$ para cualquier $E \in \Im_1 \times \Im_2$ y $y \in \mathbb{F}_2$.*

Si $E \in \Im_1 \times \Im_2$, definamos las funciones $\varphi_E : \mathbb{F}_1 \mapsto R \cup \{\infty\}$ y $\psi_E : \mathbb{F}_2 \mapsto R \cup \{\infty\}$, de la siguiente manera:

$\varphi_E(x) = \mu_2(E_x)$,

$\psi_E(y) = \mu_1(E^y)$.

PROPOSICIÓN 8.5. *Supongamos que μ_1 y μ_2 son σ-finitas, entonces φ_E y ψ_E son medibles para cualquier $E \in \Im_1 \times \Im_2$ y $\int_{\mathbb{F}_1} \varphi_E d\mu_1 = \int_{\mathbb{F}_2} \psi_E d\mu_2$.*

Demostración

Sea $\left\{F_n^{(1)}\right\}_{n\in\mathbb{N}}$ (resp. $\left\{F_n^{(2)}\right\}_{n\in\mathbb{N}}$) una familia de subconjuntos de $\mathbb{F}_1$ (resp. $\mathbb{F}_2$), ajenos por parejas y tales que $\mu_1\left(F_n^{(1)}\right) < \infty$ (resp. $\mu_2\left(F_n^{(2)}\right) < \infty$) para cualquier $n \in \mathbb{N}$ y $\mathbb{F}_1 = \cup_{n=1}^{\infty} F_n^{(1)}$ (resp. $\mathbb{F}_2 = \cup_{n=1}^{\infty} F_n^{(2)}$).

Si $E \in \Im_1 \times \Im_2$, denotemos por E^{nm} al conjunto $E \cap \left(F_n^{(1)} \times F_m^{(2)}\right)$.

Fijemos $n, m \in \mathbb{N}$ y definamos:

$\mathcal{H} = \left\{E \in \Im_1 \times \Im_2 : \ \varphi_{E^{nm}} \text{ y } \psi_{E^{nm}} \text{ son medibles y } \int_{\mathbb{F}_1} \varphi_{E^{nm}} d\mu_1 = \int_{\mathbb{F}_2} \psi_{E^{nm}} d\mu_2\right\}$.

Vamos a demostrar que $\mathcal{H}$ es una clase monótona que contiene al álgebra $\mathcal{A}$ que genera $\Im_1 \times \Im_2$.

Si $E = A \times B$ es un rectángulo medible en $\mathbb{F}_1 \times \mathbb{F}_2$, entonces:

$$\varphi_{E^{nm}}(x) = \begin{cases} \mu_2\left(B \cap F_m^{(2)}\right) & \text{si } x \in A \cap F_n^{(1)} \\ 0 & \text{en otro caso} \end{cases}$$

$$\psi_{E^{nm}}(y) = \begin{cases} \mu_1\left(A \cap F_n^{(1)}\right) & \text{si } y \in B \cap F_m^{(2)} \\ 0 & \text{en otro caso} \end{cases}$$

Así que $\varphi_{E^{nm}}$ y $\psi_{E^{nm}}$ son medible. Además:

$\int_{\mathbb{F}_1} \varphi_{E^{nm}} d\mu_1 = \mu_2\left(B \cap F_m^{(2)}\right) \mu_1\left(A \cap F_n^{(1)}\right)$

$\int_{\mathbb{F}_2} \psi_E d\mu_2 = \mu_1\left(A \cap F_n^{(1)}\right) \mu_2\left(B \cap F_m^{(2)}\right)$.

Así que $\int_{\mathbb{F}_1} \varphi_{E^{nm}} d\mu_1 = \int_{\mathbb{F}_2} \psi_{E^{nm}} d\mu_2$.

Por lo tanto, $\mathcal{H}$ contiene a cualquier rectángulo medible en $\mathbb{F}_1 \times \mathbb{F}_2$.

Sea $E_1, \ldots, E_k$ una colección finita de conjuntos en $\mathcal{H}$,ajenos por parejas y sea $E = \cup_{j=1}^{k} E_n$, entonces:

$\varphi_{E^{nm}} = \sum_{j=1}^{k} \varphi_{E_j^{nm}}$,

$\psi_{E^{nm}} = \sum_{j=1}^{k} \psi_{E_j^{nm}}$.

Así que $E \in \mathcal{H}$.

Por lo tanto, $\mathcal{H}$ contiene a la familia de conjuntos de la forma $\bigcup_{j=1}^{n} R_j$ en donde $n \in \mathbb{N}$ y $R_1, \ldots, R_n$ son rectángulos medibles en $\mathbb{F}_1 \times \mathbb{F}_2$, ajenos por parejas. Es decir, $\mathcal{A} \subset \mathcal{H}$.

Sea $\{E_k\}_{k \in \mathbb{N}}$ una sucesión no decreciente de elementos de $\mathcal{H}$ y sea $E = \cup_{k=1}^{\infty} E_k$, entonces:

$\varphi_{E^{nm}} = \lim_{k \rightsquigarrow \infty} \varphi_{E_k^{nm}}$,

$\psi_{E^{nm}} = \lim_{k \rightsquigarrow \infty} \psi_{E_k^{nm}}$.

Así que, por el teorema de la convergencia monótona, $E \in \mathcal{H}$.

Sea $\{E_k\}_{k \in \mathbb{N}}$ una sucesión no creciente de elementos de $\mathcal{H}$ y sea $E = \cap_{k=1}^{\infty} E_k$, entonces:

$\varphi_{E^{nm}} = \lim_{k \rightsquigarrow \infty} \varphi_{E_k^{nm}}$,

$\psi_{E^{nm}} = \lim_{k \rightsquigarrow \infty} \psi_{E_k^{nm}}$.

Así que, por el teorema de la convergencia dominada, $E \in \mathcal{H}$.

Utilizando el teorema de clases monótonas, concluímos entonces que $\mathcal{H} = \Im_1 \times \Im_2$.

Ahora, si $E \in \Im_1 \times \Im_2$, entonces $E = \cup_{n=1}^{\infty} \cup_{m=1}^{\infty} E^{nm}$, así que:

$\varphi_E = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \varphi_{E^{nm}}$,

$\psi_E = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \psi_{E^{nm}}$.

Así que, φ_E y ψ_E son medibles y, por el teorema de la convergencia monótona, $\int_{\mathbb{F}_1} \varphi_E d\mu_1 = \int_{\mathbb{F}_2} \psi_E d\mu_2$.

■

Proposición 8.6. *Supongamos que μ_1 y μ_2 son σ-finitas, entonces:*

$\mu_1 \times \mu_2 (E) = \int_{\mathbb{F}_1} \varphi_E d\mu_1 = \int_{\mathbb{F}_2} \psi_E d\mu_2$,

para cualquier $E \in \Im_1 \times \Im_2$.

Demostración

La función $\mu : \Im_1 \times \Im_2 \mapsto \mathbb{R} \cup \{\infty\}$ definida por:

$\mu (E) = \int_{\mathbb{F}_1} \varphi_E d\mu_1$

es una medida.

En efecto, obviamente $\mu (\emptyset) = 0$ y si $(E_n)_{n\in\mathbb{N}}$ es una sucesión de elementos de $\Im_1 \times \Im_2$, ajenos por parejas, definamos $E = \cup_{n=1}^{\infty} E_n$. Entonces, para cualquier $x \in \mathbb{F}_1$, los elementos de la sucesión $((E_n)_x)_{n\in\mathbb{N}}$ son ajenos por parejas y $E_x = \bigcup_{n=1}^{\infty} (E_n)_x$. Así que:

$\varphi_E (x) = \mu_2 (E_x) = \sum_{n=1}^{\infty} \mu_2 ((E_n)_x) = \sum_{n=1}^{\infty} \varphi_{E_n} (x)$.

Por lo tanto:

$\mu (E) = \int_{\mathbb{F}_1} \varphi_E d\mu_1 = \int_{\mathbb{F}_1} \sum_{n=1}^{\infty} \varphi_{E_n} d\mu_1 = \sum_{n=1}^{\infty} \int_{\mathbb{F}_1} \varphi_{E_n} d\mu_1 = \sum_{n=1}^{\infty} \mu (E_n)$.

Además:

Si $R = A \times B$ es un rectángulo medible en $\mathbb{F}_1 \times \mathbb{F}_2$, entonces:

$R_x = B$ para cualquier $x \in A$ y $R_x = \emptyset$ para cualquier $x \notin A$.

Así que:

$\varphi_R = \mu_2 (B) I_A$.

Por lo tanto:

$\mu (R) = \int_{\mathbb{F}_1} \varphi_R d\mu_1 = \mu_2 (B) \mu_1 (A) = \mu_1 \times \mu_2 (R)$.

Así que μ y $\mu_1 \times \mu_2$ coinciden sobre la familia de rectángulos medibles en $\mathbb{F}_1 \times \mathbb{F}_2$, los cuales forman un π-sistema de subconjuntos de $\mathbb{F}_1 \times \mathbb{F}_2$.

Por último, como μ_1 y μ_2 son σ-finitas, existen una sucesión no decreciente $(A_n)_{n\in\mathbb{N}}$ de elementos de $\Im_1$ y una sucesión no decreciente $(B_n)_{n\in\mathbb{N}}$ de elementos de $\Im_2$ tales que:

$\mathbb{F}_1 = \bigcup_{n=1}^{\infty} A_n$ y $\mu_1(A_n) < \infty$ para cualquier $n \in \mathbb{N}$,

$\mathbb{F}_2 = \bigcup_{n=1}^{\infty} B_n$ y $\mu_2(B_n) < \infty$ para cualquier $n \in \mathbb{N}$.

Así que, la sucesión de rectángulos medibles $(A_n \times B_n)_{n\in\mathbb{N}}$ es no decreciente, $\mathbb{F}_1 \times \mathbb{F}_2 = \bigcup_{n=1}^{\infty} A_n \times B_n$ y $\mu_1 \times \mu_2 (A_n \times B_n) < \infty$ para cualquier $n \in \mathbb{N}$. Por lo tanto, también se tiene $\mu(A_n \times B_n) < \infty$ para cualquier $n \in \mathbb{N}$.

Por el teorema de clases monótonas para π-sistemas, se concluye entonces que μ y $\mu_1 \times \mu_2$ coinciden sobre $\Im_1 \times \Im_2$, que es, por definición, la σ-álgebra generada por los rectángulos medibles en $\mathbb{F}_1 \times \mathbb{F}_2$.

∎

Como $\mu_1 \times \mu_2$ es una medida completa, la completación de μ es $\mu_1 \times \mu_2$. En otras palabras, μ es la restricción a $\Im_1 \times \Im_2$ de $\mu_1 \times \mu_2$.

Sabemos que todo conjunto de medida $\mu_1 \times \mu_2$ cero está contenido en un conjunto $B \in \Im_1 \times \Im_2$ de medida $\mu_1 \times \mu_2$ cero.

Sabemos también que si $f : \mathbb{F}_1 \times \mathbb{F}_2 \mapsto \overline{\mathbb{R}}$ es una función $(\Im_1 \times \Im_2)^*$-medible no negativa, entonces existe un conjunto $B \in \Im_1 \times \Im_2$ de medida $\mu_1 \times \mu_2$ cero y una función $\psi : \mathbb{F}_1 \times \mathbb{F}_2 \mapsto \overline{\mathbb{R}}$, $\Im_1 \times \Im_2$-medible no negativa, tales que $f = \psi I_{B^c} + f I_B$.

Sea $C \in (\Im_1 \times \Im_2)^*$ de medida $\mu_1 \times \mu_2$ cero y tomemos $B \in \Im_1 \times \Im_2$ de medida $\mu_1 \times \mu_2$ cero tal que $C \subset B$. Entonces:

$$\int_{\mathbb{F}_1} \left(\int_{\mathbb{F}_2} B_x(y)\, d\mu_2(y) \right) d\mu_1(x) = \int_{\mathbb{F}_1 \times \mathbb{F}_2} I_B d\mu = 0.$$

Así que si definimos: $E_1 = \{x \in \mathbb{F}_1 : \mu_2(B_x) > 0\}$, entonces $E_1 \in \Im_1$ y $\mu_1(E_1) = 0$.

Si $x \notin E_1$, entonces $\mu_2(B_x) = 0$. Además, $C_x \subset B_x$, así que $C_x \in \Im_2$ y $\mu_2(C_x) = 0$.

Por lo tanto, para casi toda x, $C_x \in \Im_1$ y $\mu_2(C_x) = 0$.

De la misma manera, para casi toda y, $C^y \in \Im_2$ y $\mu_1(C^y) = 0$.

Si $f : \mathbb{F}_1 \times \mathbb{F}_2 \mapsto \mathbb{R} \cup \{\infty\}$ es una función $\Im_1 \times \Im_2$-medible no negativa, definamos, para cada $x \in \mathbb{F}_1$, la función $f_x : \mathbb{F}_2 \mapsto \mathbb{R} \cup \{\infty\}$, de la siguiente manera:

$$f_x(y) = f(x, y)$$

y, para cada $y \in \mathbb{F}_2$, la función $f^y : \mathbb{F}_1 \mapsto \mathbb{R} \cup \{\infty\}$, de la siguiente manera:

$$f^y(x) = f(x, y).$$

Proposición 8.7. *Para cualquier función $\Im_1 \times \Im_2$-medible no negativa, $f : \mathbb{F}_1 \times \mathbb{F}_2 \mapsto \mathbb{R} \cup \{\infty\}$, y cualquier $x \in \mathbb{F}_1$, la función f_x es $\Im_2$-medible.*

Demostración

Sea B un conjunto boreliano en $\mathbb{R}\cup\{\infty\}$ y $E = f^{-1}(B)$. Entonces:

$f_x^{-1}(B) = \{y \in \mathbb{F}_2 : f(x,y) \in B\} = \{y \in \mathbb{F}_2 : (x,y) \in E\} = E_x.$

$\blacksquare$

COROLARIO 8.7. *Para cualquier función $\Im_1\times\Im_2$-medible no negativa, $f : \mathbb{F}_1\times\mathbb{F}_2 \mapsto \mathbb{R}\cup\{\infty\}$, y cualquier $y \in \mathbb{F}_2$, la función f^y es $\Im_1$-medible.*

PROPOSICIÓN 8.8. *Supongamos que μ_1 y μ_2 son σ-finitas, entonces las funciones $x \mapsto \int_{\mathbb{F}_2} f_x d\mu_2(y)$ y $y \mapsto \int_{\mathbb{F}_1} f^y d\mu_1(x)$ son medibles para cualquier función no negativa $\Im_1 \times \Im_2$-medible, $f : \mathbb{F}_1 \times \mathbb{F}_2 \mapsto \mathbb{R}\cup\{\infty\}$, y:*

$$\int_{\mathbb{F}_1}\left(\int_{\mathbb{F}_2} f_x d\mu_2(y)\right) d\mu_1(x) = \int_{\mathbb{F}_2}\left(\int_{\mathbb{F}_1} f^y d\mu_1(x)\right) d\mu_2(y) = \int_{\mathbb{F}_1\times\mathbb{F}_2} f d(\mu_1\times\mu_2).$$

Demostración

Si $f = I_E$, con $E \in \Im_1 \times \Im_2$, entonces las funciones $x \mapsto \int_{\mathbb{F}_2} f_x d\mu_2(y) = \mu_2(E_x) = \varphi_E(x)$ y $y \mapsto \int_{\mathbb{F}_1} f^y d\mu_1(x) = \mu_1(E^y) = \psi_E(y)$ son medibles y:

$$\int_{\mathbb{F}_1}\left(\int_{\mathbb{F}_2} f_x d\mu_2(y)\right) d\mu_1(x) = \int_{\mathbb{F}_1} \varphi_E d\mu_1 = \mu_1\times\mu_2(E) = \int_{\mathbb{F}_1\times\mathbb{F}_2} f d(\mu_1\times\mu_2),$$

$$\int_{\mathbb{F}_2}\left(\int_{\mathbb{F}_1} f^y d\mu_1(x)\right) d\mu_2(y) = \int_{\mathbb{F}_2} \psi_E d\mu_2 = \mu_1\times\mu_2(E) = \int_{\mathbb{F}_1\times\mathbb{F}_2} f d(\mu_1\times\mu_2).$$

Si f es una función medible simple no negativa, digamos $f = \sum_{k=1}^m b_k I_{E_k}$, donde $b_1,\ldots,b_m \in \mathbb{R}$ y $E_1,\ldots,E_m \in \Im_1 \times \Im_2$, entonces $\int_{\mathbb{F}_2} f_x d\mu_2(y) = \sum_{k=1}^m b_k\mu_2\left((E_k)_x\right)$, así que la función $x \mapsto \int_{\mathbb{F}_2} f_x d\mu_2(y)$ es medible. Además:

$$\int_{\mathbb{F}_1}\left(\int_{\mathbb{F}_2} f_x d\mu_2(y)\right) d\mu_1(x) = \int_{\mathbb{F}_1}\left(\sum_{k=1}^m b_k\mu_2\left((E_k)_x\right)\right) d\mu_1(x)$$

$$= \sum_{k=1}^m b_k \int_{\mathbb{F}_1} \varphi_{E_k} d\mu_1 = \sum_{k=1}^m b_k\mu_1\times\mu_2(E_k) = \int_{\mathbb{F}_1\times\mathbb{F}_2} f d(\mu_1\times\mu_2).$$

De la misma manera, se tiene $\int_{\mathbb{F}_2}\left(\int_{\mathbb{F}_1} f^y d\mu_1(x)\right) d\mu_2(y) = \int_{\mathbb{F}_1\times\mathbb{F}_2} f d(\mu_1\times\mu_2)$.

Si f es cualquier medible, sea $(\varphi_n)_{n\in\mathbb{N}}$ una sucesión no decreciente de funciones medibles simples no negativas tal que $f = \lim_{n\rightsquigarrow\infty}\varphi_n$. Entonces:

$$\int_{\mathbb{F}_2} f_x d\mu_2(y) = \int_{\mathbb{F}_2} \lim_{n\rightsquigarrow\infty}(\varphi_n)_x d\mu_2(y) = \lim_{n\rightsquigarrow\infty}\int_{\mathbb{F}_2}(\varphi_n)_x d\mu_2(y).$$

Así que la función $x \mapsto \int_{\mathbb{F}_2} f_x d\mu_2(y)$ es medible. Además:

$$\int_{\mathbb{F}_1}\left(\int_{\mathbb{F}_2} f_x d\mu_2(y)\right) d\mu_1(x) = \int_{\mathbb{F}_1}\left(\lim_{n\rightsquigarrow\infty}\int_{\mathbb{F}_2}(\varphi_n)_x d\mu_2(y)\right) d\mu_1(x)$$

$= \text{lím}_{n \rightsquigarrow \infty} \int_{\mathbb{F}_1} \left(\int_{\mathbb{F}_2} (\varphi_n)_x \, d\mu_2 (y) \right) d\mu_1 (x)$

$= \text{lím}_{n \rightsquigarrow \infty} \int_{\mathbb{F}_1 \times \mathbb{F}_2} \varphi_n d (\mu_1 \times \mu_2) = \int_{\mathbb{F}_1 \times \mathbb{F}_2} f d (\mu_1 \times \mu_2)$.

De la misma manera, se tiene $\int_{\mathbb{F}_2} \left(\int_{\mathbb{F}_1} f^y d\mu_1 (x) \right) d\mu_2 (y) = \int_{\mathbb{F}_1 \times \mathbb{F}_2} f d (\mu_1 \times \mu_2)$.

■

TEOREMA 8.12 (**Teorema de Tonelli**). *Supongamos que μ_1 y μ_2 son medidas completas y σ-finitas. Sea $f : \mathbb{F}_1 \times \mathbb{F}_2 \mapsto \mathbb{R} \cup \{\infty\}$ una función $(\Im_1 \times \Im_2)^*$-medible no negativa y definamos $C_1 = \{x \in \mathbb{F}_1 : f_x$ es $\Im_2$-medible$\}$, $C_2 = \{y \in \mathbb{F}_2 : f^y$ es $\Im_1$-medible$\}$. Entonces:*

(i) *C_1^c y C_2^c tienen medida cero.*
(ii) *Las funciones $x \to I_{C_1} (x) \int_{\mathbb{F}_2} f_x d\mu_2$ y $y \to I_{C_2} (y) \int_{\mathbb{F}_1} f^y d\mu_1$ son medibles.*
(iii) $\int_{C_1} \left(\int_{\mathbb{F}_2} f_x d\mu_2 (y) \right) d\mu_1 (x) = \int_{C_2} \left(\int_{\mathbb{F}_1} f^y d\mu_1 (x) \right) d\mu_2 (y)$
$= \int_{\mathbb{F}_1 \times \mathbb{F}_2} f d (\mu_1 \times \mu_2)^*$.

Demostración

Sean $\psi : \mathbb{F}_1 \times \mathbb{F}_2 \mapsto \overline{\mathbb{R}}$ una función $\Im_1 \times \Im_2$-medible no negativa y $B \in \Im_1 \times \Im_2$ de medida $(\mu_1 \times \mu_2)^*$ cero tales que $f = \psi I_{B^c} + f I_B$.

Definamos $E_1 = \{x \in \mathbb{F}_1 : \mu_2 (B_x) > 0\}$. Entonces $E_1 \in \Im_1$ y:

$\mu_1 (E_1) = 0$.

Si $x \notin E_1$, entonces $\mu_2 (B_x) = 0$. Así que $f_x = \psi_x$ para casi toda $y \in \mathbb{F}_2$. Por lo tanto, f_x es $\Im_2$-medible y .

Así que $E_1^c \subset C_1$ y entonces $C_1^c \subset E_1$. Así que $C_1^c \in \Im_1$ y $\mu_1 (C_1^c) = 0$.

También, como $E_1^c \subset C_1$, entonces $C_1 = E_1^c \cup (C_1 \cap E_1)$, así que, para cualquier $x \in C_1$, se tiene:

$\int_{\mathbb{F}_2} f_x d\mu_2 = I_{E_1^c} (x) \int_{\mathbb{F}_2} f_x d\mu_2 + I_{C_1 \cap E_1} (x) \int_{\mathbb{F}_2} f_x d\mu_2$

$= I_{E_1^c} (x) \int_{\mathbb{F}_2} \psi_x d\mu_2 + I_{C_1 \cap E_1} (x) \int_{\mathbb{F}_2} f_x d\mu_2$.

Por lo tanto, $I_{C_1} (x) \int_{\mathbb{F}_2} f_x d\mu_2 = I_{E_1^c} (x) \int_{\mathbb{F}_2} \psi_x d\mu_2$ para casi toda $x \in \mathbb{F}_1$. Así que la función $x \to I_{C_1} (x) \int_{\mathbb{F}_2} f_x d\mu_2$ es medible y:

$\int_{C_1} \left(\int_{\mathbb{F}_2} f_x d\mu_2 (y) \right) d\mu_1 (x) = \int_{\mathbb{F}_1} I_{C_1} (x) \left(\int_{\mathbb{F}_2} f_x d\mu_2 (y) \right) d\mu_1 (x)$

$= \int_{\mathbb{F}_1} I_{E_1^c} (x) \left(\int_{\mathbb{F}_2} \psi_x d\mu_2 (y) \right) d\mu_1 (x) = \int_{\mathbb{F}_1} \left(\int_{\mathbb{F}_2} \psi_x d\mu_2 (y) \right) d\mu_1 (x)$

$= \int_{\mathbb{F}_1 \times \mathbb{F}_2} \psi d (\mu_1 \times \mu_2) = \int_{\mathbb{F}_1 \times \mathbb{F}_2} f d (\mu_1 \times \mu_2)^*$.

De la misma manera, si definimos $E_2 = \{y \in \mathbb{F}_2 : \mu_1 (B^y) > 0\}$, se tiene $E_2 \in \Im_2$, $\mu_2 (E_2) = 0$, $E_2^c \subset C_2$, $C_2^c \in \Im_2$ y $\mu_2 (C_2^c) = 0$, la función $y \to I_{C_2} (y) \int_{\mathbb{F}_1} f^y d\mu_1 = I_{E_2^c} (y) \int_{\mathbb{F}_1} \psi^y d\mu_1$ es medible y:

$$\int_{C_2} \left(\int_{\mathbb{F}_1} f^y d\mu_1 (x) \right) d\mu_2 (y) = \int_{\mathbb{F}_2} I_{C_2} (y) \left(\int_{\mathbb{F}_1} f^y d\mu_1 (x) \right) d\mu_2 (y)$$

$$= \int_{\mathbb{F}_2} I_{E_2^c} (x) \left(\int_{\mathbb{F}1} \psi^y d\mu_1 (x) \right) d\mu_2 (y) = \int_{\mathbb{F}_2} \left(\int_{\mathbb{F}_1} \psi^y d\mu_1 (x) \right) d\mu_2 (y)$$

$$= \int_{\mathbb{F}_1 \times \mathbb{F}_2} \psi d (\mu_1 \times \mu_2) = \int_{\mathbb{F}_1 \times \mathbb{F}_2} f d (\mu_1 \times \mu_2)^*.$$

■

TEOREMA 8.13 (**Teorema de Fubini**). *Supongamos que μ_1 y μ_2 son medidas completas y σ-finitas, sea $f : \mathbb{F}_1 \times \mathbb{F}_2 \mapsto \mathbb{R} \cup \{-\infty, \infty\}$ una función $\Im_1 \times \Im_2$-medible (o $(\Im_1 \times \Im_2)^*$-medible) e integrable y definamos $C_1 = \{x \in \mathbb{F}_1 : f_x \text{ es } \mu_2\text{-integrable}\}$, $C_2 = \{y \in \mathbb{F}_2 : f^y \text{ es } \mu_1\text{-integrable}\}$. Entonces:*

(i) *C_1^c y C_2^c tienen medida cero.*
(ii) *Las funciones $x \to I_{C_1} (x) \int_{\mathbb{F}_2} f_x d\mu_2$ y $y \to I_{C_2} (y) \int_{\mathbb{F}_1} f^y d\mu_1$ son integrables.*
(iii) $\int_{C_1} \left(\int_{\mathbb{F}_2} f_x d\mu_2 (y) \right) d\mu_1 (x) = \int_{C_2} \left(\int_{\mathbb{F}_1} f^y d\mu_1 (x) \right) d\mu_2 (y)$
$= \int_{\mathbb{F}_1 \times \mathbb{F}_2} f d (\mu_1 \times \mu_2)^*$.

Demostración

Como f es integrable, entonces f^+ es también integrable, así que:

$$\int_{\mathbb{F}_1} \left(\int_{\mathbb{F}_2} f^+ (x, y) d\mu_2 (y) \right) d\mu_1 (x) = \int_{\mathbb{F}_1 \times \mathbb{F}_2} f^+ d (\mu_1 \times \mu_2)^* < \infty.$$

Por lo tanto, $\int_{\mathbb{F}_2} f^+ (x, y) d\mu_2 (y) < \infty$ para casi toda $x \in \mathbb{F}_1$.

De la misma manera, $\int_{\mathbb{F}_2} f^- (x, y) d\mu_2 (y) < \infty$ para casi toda $x \in \mathbb{F}_1$.

Por lo tanto, f_x es μ_2-integrable para casi toda $x \in \mathbb{F}_1$.

Además:

$$\int_{C_1} \left(\int_{\mathbb{F}_2} f^+ (x, y) d\mu_2 (y) \right) d\mu_1 (x) = \int_{\mathbb{F}_1} \left(\int_{\mathbb{F}_2} f^+ (x, y) d\mu_2 (y) \right) d\mu_1 (x) < \infty,$$

$$\int_{C_1} \left(\int_{\mathbb{F}_2} f^- (x, y) d\mu_2 (y) \right) d\mu_1 (x) = \int_{\mathbb{F}_1} \left(\int_{\mathbb{F}_2} f^- (x, y) d\mu_2 (y) \right) d\mu_1 (x) < \infty.$$

Así que la función:

$$x \to I_{C_1} (x) \int_{\mathbb{F}_2} f_x d\mu_2$$

$$= I_{C_1} (x) \int_{\mathbb{F}_2} f^+ (x, y) d\mu_2 (y) - I_{C_1} (x) \int_{\mathbb{F}_2} f^- (x, y) d\mu_2 (y)$$

es integrable.

Por último:

$\int_{C_1}\left(\int_{\mathbb{F}_2} f_x d\mu_2\right) d\mu_1$

$= \int_{C_1}\left(\int_{\mathbb{F}_2} f^+(x,y)\, d\mu_2(y)\right) d\mu_1(x) - \int_{C_1}\left(\int_{\mathbb{F}_2} f^-(x,y)\, d\mu_2(y)\right) d\mu_1(x)$

$= \int_{\mathbb{F}_1\times\mathbb{F}_2} f^+ d(\mu_1\times\mu_2)^* - \int_{\mathbb{F}_1\times\mathbb{F}_2} f^{\cdot -} d(\mu_1\times\mu_2)^* = \int_{\mathbb{F}_1\times\mathbb{F}_2} f^{\cdot} d(\mu_1\times\mu_2)^*.$

■

8.4. Proyección de medidas

TEOREMA 8.14. *Sean $(\mathbb{G},\mathcal{G})$ un espacio medible y $h:(\mathbb{F},\Im,\mu)\mapsto(\mathbb{G},\mathcal{G})$ una función medible. Entonces la función $\nu:\mathcal{G}\mapsto\overline{\mathbb{R}}$ definida por $\nu(B)=\mu(\{y\in\mathbb{F}:h(y)\in B\})$ es una medida.*

Demostración

Obviamente ν es no negativa y $\nu(\emptyset)=0$. Para demostrar que ν es σ-aditiva, sea $B_1, B_2, \ldots$ una familia numerable de elementos de $\mathcal{G}$ tal que $B_i\cap B_j=\emptyset$ para $i\neq j$, entonces, definiendo, para cada $i\in\mathbb{N}$, $A_i=\{y\in\mathbb{F}:h(y)\in B_i\}$, los conjuntos $A_1, A_2, \ldots$ son ajenos por parejas, así que:

$\nu\left(\bigcup_{i=1}^{\infty} B_i\right) = \mu\left(\left\{y\in\mathbb{F}:h(y)\in\bigcup_{i=1}^{\infty} B_i\right\}\right) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty}\mu(A_i) = \sum_{i=1}^{\infty}\nu(B_i).$

■

DEFINICIÓN 8.4. *A la medida ν definida como en el teorema anterior se le llama la proyección de μ bajo la función h.*

TEOREMA 8.15. *Sea $h:(\mathbb{F},\Im,\mu)\mapsto(\mathbb{G},\mathcal{G})$ una función medible. Entonces si $\nu:\mathcal{G}\mapsto\overline{\mathbb{R}}$ es la proyección de μ bajo h y $f:\mathbb{G}\mapsto\overline{\mathbb{R}}$ es una función medible no negativa, se tiene:*

$\int_{\mathbb{G}} f d\nu = \int_{\mathbb{F}} f\circ h d\mu.$

Demostración

Consideremos primero una función simple no negativa $\varphi:\mathbb{G}\mapsto\mathbb{R}$, con representación canónica $\varphi=\sum_{j=1}^{m} b_j I_{E_j}$.

Entonces:

$\varphi\circ h = \sum_{j=1}^{m} b_j I_{E_j}\circ h = \sum_{j=1}^{m} b_j I_{h^{-1}(E_j)},$

$\int_{\mathbb{G}}\varphi d\nu = \sum_{j=1}^{m} b_j\nu(E_j) = \sum_{j=1}^{m} b_j\mu(h^{-1}(E_j)) = \int_{\mathbb{F}}\varphi\circ h d\mu.$

Consideremos ahora una sucesión no decreciente $(\varphi_n)_{n\in\mathbb{N}}$ de funciones simples no negativas tales que $\text{lím}_{n\rightsquigarrow\infty}\varphi_n(y) = f(y)$ para cualquier $y \in \mathbb{G}$. Entonces, para cualquier $n \in \mathbb{N}$ la función $\varphi_n \circ h$ es no negativa, la sucesión $(\varphi_n \circ h)_{n\in\mathbb{N}}$ es no decreciente y $\text{lím}_{n\rightsquigarrow\infty}\varphi_n \circ h(x) = f \circ h(x)$ para cualquier $x \in \mathbb{F}$.; así que:

$\int_{\mathbb{G}} f d\nu = \text{lím}_{n\rightsquigarrow\infty}\int_{\mathbb{G}} \varphi_n d\nu = \text{lím}_{n\rightsquigarrow\infty}\int_{\mathbb{F}} \varphi_n \circ h d\mu = \int_{\mathbb{F}} f \circ h d\mu.$

■

COROLARIO 8.8. *Sea $h : (\mathbb{F}, \Im, \mu) \mapsto (\mathbb{G}, \mathcal{G})$ una función medible. Entonces si $\nu : \mathcal{G} \mapsto \overline{\mathbb{R}}$ es la proyección de μ bajo h y $f : \mathbb{G} \mapsto \overline{\mathbb{R}}$ es una función integrable, se tiene:*

$\int_{\mathbb{G}} f d\nu = \int_{\mathbb{F}} f \circ h d\mu.$

Demostración

Por la proposición anterior, se tiene:

$\int_{\mathbb{F}} f^+ \circ h d\mu = \int_{\mathbb{G}} f^+ d\nu < \infty,$

$\int_{\mathbb{F}} f^- \circ h d\mu = \int_{\mathbb{G}} f^- d\nu < \infty.$

Así que:

$\int_{\mathbb{G}} f d\nu = \int_{\mathbb{G}} f^+ d\nu - \int_{\mathbb{G}} f^- d\nu = \int_{\mathbb{F}} f^+ \circ h d\mu - \int_{\mathbb{F}} f^- \circ h d\mu = \int_{\mathbb{F}} f \circ h d\mu.$

■

A continuación vamos a presentar un caso particular de los resultados anteriores y que es de especial importancia ya que muestra que, para cualquier cualquier medida ν, sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, que asigne un valor finito a cualquier conjunto boreliano acotado, existe una función c que proyecta la medida de Lebesgue en ν, lo cual tiene como corolario que las integrales con respecto a ν se pueden expresar como integrales con respecto a la medida de Lebesgue.

Sea ν_0 una medida, no idénticamente cero, sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ tal que los intervalos acotados tienen medida finita y sea $F_{\nu_0} : \mathbb{R} \mapsto \mathbb{R}$ una función no decreciente y continua por la derecha tal que $\nu_0((a, b]) = F_{\nu_0}(b) - F_{\nu_0}(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$.

F_{ν_0} puede ser constante en uno o más intervalosde la forma $[a_0, b_0)$, donde $a_0, b_0 \in \mathbb{R}$, tales que $F_{\nu_0}(x) < F_{\nu_0}(a_0)$ para cualquier $x < a_0$ y $F_{\nu_0}(x) > F_{\nu_0}(a_0)$ para cualquier $x > b_0$. El conjunto de intervalos de este tipo es a lo más infinito numerable, así que los podemos denotar por $I_1 = [a_1, b_1), I_2 = [a_2, b_2), \ldots$. También puede ser constante en un intervalo $I_0 = (a_0, b_0)$, donde $a_0 = -\infty$ y $b_0 \in \mathbb{R}$, tal que $F_{\nu_0}(x) > F_{\nu_0}(b_0-)$ para cualquier $x > b_0$, o en un intervalo $I_\infty = [a_\infty, b_\infty)$, donde $a_\infty \in \mathbb{R}$ y $b_\infty = \infty$, tal que $F_{\nu_0}(x) < F_{\nu_0}(a_\infty)$ para cualquier $x < a_0$. Denotaremos por $\mathcal{H}$ a la familia formada por todos los intervalos de los tipos mencionados y por K a la unión de todos ellos. También, denotaremos por s_k al valor que toma F_{ν_0} en el intervalo I_k.

F_{ν_0} puede ser continua en uno o varios puntos que no pertenecen a K. Denotaremos por C a ese conjunto de puntos.

F_{ν_0} puede ser discontinua en uno o varios puntos que no pertenecen a K, los cuales forman un conjunto a lo más infinito numerable. Denotaremos por D a ese conjunto de puntos.

Obviamente los conjuntos K, C y D son ajenos por parejas y su unión en $\mathbb{R}$.

Sean $m = \lim_{x\to-\infty} F_{\nu_0}(x)$ y $M = \lim_{x\rightsquigarrow\infty} F_{\nu_0}(x)$ y definamos la función $c : (m, M) \mapsto \mathbb{R}$ mediante la relación $c(t) = \inf\{x \in \mathbb{R} : F_{\nu_0}(x) > t\}$.

Evidentemente, c es una función no decreciente y, para cualquier intervalo $I_k \in \mathcal{H}$, no existe $t \in (m, M)$ tal que $c(t) \in (a_k, b_k)$. Además, considerando a c como función de $((m, M), \mathfrak{B}((m, M)))$ en $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, c es medible ya que es no decreciente.

Se tienen las siguientes 5 situaciones:

1. Si $x_0 \in I_k \in \mathcal{H}$, donde $k \notin \{0, \infty\}$, se tiene:

$c(s_k-) = a_k$ y $c(s_k) = b_k$. Por lo tanto:

$\{y \in (m, M) : c(y) \leq x_0\} = \{y \in (m, M) : c(y) \leq a_k\} = (m, s_k)$,

$\{y \in (m, M) : c(y) > x_0\} = \{y \in (m, M) : c(y) \geq b_k\} = [s_k, M)$.

2. Si $x_0 \in (-\infty, b_0) \in \mathcal{H}$, entonces $s_0 = m \in \mathbb{R}$ y $c(s_0+) = b_0$. Por lo tanto:

$\{y \in (m, M) : c(y) \leq x_0\} = \emptyset$,

$\{y \in (m, M) : c(y) > x_0\} = \{y \in (m, M) : c(y) \geq b_0\} = (m, M)$.

3. Si $x_0 \in [a_\infty, \infty) \in \mathcal{H}$, entonces $s_\infty = M \in \mathbb{R}$ y $c(s_\infty-) = a_\infty$. Por lo tanto:

$\{y \in (m, M) : c(y) \leq x_0\} = \{y \in (m, M) : c(y) \leq a_\infty\} = (m, M)$,

$\{y \in (m, M) : c(y) > x_0\} = \emptyset$.

4. Si $x_0 \in C$ y $F_{\nu_0}(x_0) = t_0$, entonces $c(t_0) = x_0$, $c(t) < x_0$ para cualquier $t < t_0$ y $c(t) > x_0$ para cualquier $t > t_0$. Por lo tanto:

$\{y \in (m, M) : c(y) \leq x_0\} = (m, t_0]$,

$\{y \in (m, M) : c(y) > x_0\} = (t_0, M)$.

5. Si $x_0 \in D$, sea $t_1 = F_{\nu_0}(x_0-)$ y $t_2 = F_{\nu_0}(x_0)$; entonces $c(t) = x_0$ para cualquier $t \in [t_1, t_2]$, $c(t) < x_0$ para cualquier $t < t_1$ y $c(t) > x_0$ para cualquier $t > t_2$. Por lo tanto:

$\{y \in (m, M) : c(y) \leq x_0\} = (m, t_2]$,

$\{y \in (m, M) : c(y) > x_0\} = (t_2, M)$.

TEOREMA 1. *Sean ν_0 una medida sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ tal que los intervalos acotados tienen medida finita, $F_{\nu_0} : \mathbb{R} \mapsto \mathbb{R}$ una función no decreciente y continua por la derecha tal que $\nu_0((a,b]) = F_{\nu_0}(b) - F_{\nu_0}(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$, $m = \lim_{x\to-\infty} F_{\nu_0}(x)$, $M = \lim_{x\rightsquigarrow\infty} F_{\nu_0}(x)$, λ la medida de Lebesgue en el intervalo (m, M) y $c : (m, M) \mapsto \mathbb{R}$ definida por $c(t) = \inf\{x \in \mathbb{R} : F_{\nu_0}(x) > t\}$. Entonces, considerando a c como función de $((m, M), \mathfrak{B}((m, M)))$ en $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, la proyección de λ bajo c es ν_0.*

Demostración

Sea ν la proyección de λ bajo c, es decir:

$\nu(B) = \lambda(c^{-1}(B))$ para cualquier conjunto $B \in \mathfrak{B}(\mathbb{R})$.

En particular:

$\nu((a,b]) = \lambda(\{y \in (m, M) : c(y) \in (a,b]\})$ para cualquier pareja de números reales, a y b, tales que $a < b$.

Sean $a, b \in \mathbb{R}$ tales que $a < b$. Tenemos entonces varias posibilidades:

1. a y b pertenecen a algún intervalo $I \in \mathcal{H}$.

En este caso, si J es el interior de I, no existe $t \in (m, M)$ tal que $c(t) \in J$. Entonces, como $(a,b] \subset J$, se tiene:

$\{y \in (m, M) : c(y) \in (a,b]\} = \emptyset$.

Por lo tanto:

$\nu((a,b]) = \lambda(\{y \in (m, M) : c(y) \in (a,b]\}) = 0 = F_{\nu_0}(b) - F_{\nu_0}(a) = \nu_0((a,b])$.

2. a pertenece a algún intervalo $I_k \in \mathcal{H}$ y $b \in C$.

En este caso, $k \neq \infty$, ya que $b > a$, y $\{y \in (m, M) : c(y) \leq b\} = (m, t_0]$, donde $t_0 = F_{\nu_0}(b)$.

Si $k \neq 0$, $\{y \in (m, M) : c(y) > a\} = [s_k, M)$, así que:

$\{y \in (m, M) : c(y) \in (a,b]\} = [s_k, t_0]$.

Si $k = 0$, $\{y \in (m, M) : c(y) > a\} = (m, M)$, así que:

$\{y \in (m, M) : c(y) \in (a,b]\} = (s_k, t_0]$.

Por lo tanto:

$\nu((a,b]) = \lambda(\{y \in (m, M) : c(y) \in (a,b]\}) = t_0 - s_k = F_{\nu_0}(b) - F_{\nu_0}(a) = \nu_0((a,b])$.

3. a pertenece a algún intervalo I_k y $b \in D$.

En este caso, $k \neq \infty$, ya que $b > a$, y $\{y \in (m, M) : c(y) \leq b\} = (m, t_2]$, donde $t_2 = F_{\nu_0}(b)$.

Si $k \neq 0$, $\{y \in (m, M) : c(y) > a\} = [s_k, M)$, así que:

$\{y \in (m, M) : c(y) \in (a, b]\} = [s_k, t_2]$.

Si $k = 0$, $\{y \in (m, M) : c(y) > a\} = (m, M)$, así que:

$\{y \in (m, M) : c(y) \in (a, b]\} = (s_k, t_2]$.

Por lo tanto:

$\nu((a, b]) = \lambda(\{y \in (m, M) : c(y) \in (a, b]\}) = t_2 - s_k = F_{\nu_0}(b) - F_{\nu_0}(a) = \nu_0((a, b])$.

4. $a \in C$ y b pertenece a algún intervalo I_k.

En este caso, $k \neq 0$, ya que $a < b$, y $\{y \in (m, M) : c(y) > a\} = (t_0, M)$, donde $t_0 = F_{\nu_0}(a)$.

Si $k \neq \infty$, $\{y \in (m, M) : c(y) \leq b\} = (m, s_k)$, así que:

$\{y \in (m, M) : c(y) \in (a, b]\} = (t_0, s_k)$.

Si $k = \infty$, $\{y \in (m, M) : c(y) \leq b\} = (m, M)$, así que:

$\{y \in (m, M) : c(y) \in (a, b]\} = (t_0, s_k)$.

Por lo tanto:

$\nu((a, b]) = \lambda(\{y \in (m, M) : c(y) \in (a, b]\}) = s_k - t_0 = F_{\nu_0}(b) - F_{\nu_0}(a) = \nu_0((a, b])$.

5. $a, b \in C$.

En este caso:

$\{y \in (m, M) : c(y) > a\} = (t_1, M)$, donde $t_1 = F_{\nu_0}(a)$,

$\{y \in (m, M) : c(y) \leq b\} = (m, t_2]$, donde $t_2 = F_{\nu_0}(b)$.

Así que:

$\{y \in (m, M) : c(y) \in (a, b]\} = (t_1, t_2]$.

Por lo tanto:

$\nu((a, b]) = \lambda(\{y \in (m, M) : c(y) \in (a, b]\}) = t_2 - t_1 = F_{\nu_0}(b) - F_{\nu_0}(a) = \nu_0((a, b])$.

6. $a \in C$ y $b \in D$.

En este caso:

$\{y \in (m, M) : c(y) > a\} = (t_0, M)$, donde $t_0 = F_{\nu_0}(x_0)$,

$\{y \in (m, M) : c(y) \leq b\} = (m, t_2]$, donde $t_2 = F_{\nu_0}(b)$.

Así que:

$\{y \in (m, M) : c(y) \in (a, b]\} = (t_0, t_2]$.

Por lo tanto:

$\nu((a, b]) = \lambda(\{y \in (m, M) : c(y) \in (a, b]\}) = t_2 - t_0 = F_{\nu_0}(b) - F_{\nu_0}(a) = \nu_0((a, b])$.

7. $a \in D$ y b pertenece a algún intervalo I_k.

En este caso, $k \neq 0$ ya que $a < b$, y $\{y \in (m, M) : c(y) > a\} = (t_2, M)$, donde $t_2 = F_{\nu_0}(a)$.

Si $k \neq \infty$, $\{y \in (m, M) : c(y) \leq b\} = (m, s_k)$, así que:

$\{y \in (m, M) : c(y) \in (a, b]\} = (t_2, s_k)$.

Si $k = \infty$, $\{y \in (m, M) : c(y) \leq b\} = (m, M)$, así que:

$\{y \in (m, M) : c(y) \in (a, b]\} = (t_2, s_k)$.

Por lo tanto:

$\nu((a, b]) = \lambda(\{y \in (m, M) : c(y) \in (a, b]\}) = s_k - t_2 = F_{\nu_0}(b) - F_{\nu_0}(a) = \nu_0((a, b])$.

8. $a \in D$ y $b \in C$.

En este caso:

$\{y \in (m, M) : c(y) > a\} = (t_2, M)$, donde $t_2 = F_{\nu_0}(a)$,

$\{y \in (m, M) : c(y) \leq b\} = (m, t_0]$, donde $t_0 = F_{\nu_0}(b)$.

Así que:

$\{y \in (m, M) : c(y) \in (a, b]\} = (t_2, t_0]$.

Por lo tanto:

$\nu((a, b]) = \lambda(\{y \in (m, M) : c(y) \in (a, b]\}) = t_0 - t_2 = F_{\nu_0}(b) - F_{\nu_0}(a) = \nu_0((a, b])$.

9. $a, b \in D$.

En este caso:

$\{y \in (m, M) : c(y) > a\} = (t_2, M)$, donde $t_2 = F_{\nu_0}(a)$,

$\{y \in (m, M) : c(y) \leq b\} = (m, t_2']$, donde $t_2' = F_{\nu_0}(b)$.

Así que:

$\{y \in (m, M) : c(y) \in (a, b]\} = (t_2, t_2']$.

Por lo tanto:

$\nu((a,b]) = \lambda(\{y \in \mathbb{R} : c(y) \in (a,b]\}) = t_2' - t_2 = F_{\nu_0}(b) - F_{\nu_0}(a) = \nu_0((a,b])$.

Así que, en cualquier caso, $\nu((a,b]) = \nu_0((a,b])$ para cualquier pareja de números reales, a y b, tales que $a < b$.

Aplicando el teorema de clases monótonas, concluimos que $\nu(B) = \nu_0(B)$ para cualquier conjunto $B \in \mathfrak{B}(\mathbb{R})$.

■

COROLARIO 8.9. *Sean ν una medida sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ tal que los intervalos acotados tienen medida finita, $F_\nu : \mathbb{R} \mapsto \mathbb{R}$ una función no decreciente y continua por la derecha tal que $\nu((a,b]) = F_\nu(b) - F_\nu(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$, $m = \text{lím}_{x\to-\infty} F_\nu(x)$, $M = \text{lím}_{x\rightsquigarrow\infty} F_\nu(x)$, λ la medida de Lebesgue en el intervalo (m, M) y $c : (m, M) \mapsto \mathbb{R}$ definida por $c(t) = \text{ínf}\{x \in \mathbb{R} : F_\nu(x) > t\}$. Entonces, si $f : (\mathbb{R}, \mathfrak{B}(\mathbb{R}), \nu) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una función medible, no negativa o integrable, se tiene:*

$\int_{\mathbb{R}} f d\nu = \int_m^M (f \circ c)\, d\lambda$.

CAPÍTULO 9

LA INTEGRAL DE LEBESGUE STIELTJES

La integral de Lebesgue-Stieltjes es la que se se obtiene al considerar la integral con respecto a una medida generada por una función de variación acotada definida sobre los borelianos de $\mathbb{R}$. Así que tiene las propiedades de la integral expuestas en el capítulo 6. Además, es una de las componentes de la integral estocástica, cuya definición y estudio es el objetivo de este trabajo. De aquí la importancia de estudiar con detalle las propiedades particulares de una integral de este tipo.

Para los fines que requerimos en el estudio de la integración estocástica, una función de variación acotada será una función del tiempo; es por eso que en este capítulo vamos a considerar funciones no decrecientes o de variación acotada definidas sobre $\mathbb{R}^{+}$.

Sea $F : \mathbb{R} \to \mathbb{R}$ una función no decreciente. De acuerdo con el teorema 6.3, F genera una medida μ_F definida sobre $\mathfrak{B}(\mathbb{R})$. Completemos el espacio de medida $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \mu_F)$ y denotemos por $\mathfrak{B}_F(\mathbb{R})$ a la σ-álgebra de los borelianos completada.

DEFINICIÓN 9.1. *Si $f : (\mathbb{R}, \mathfrak{B}_F(\mathbb{R}), \mu_F) \to \mathbb{R}$ es una función medible e integrable, diremos que f es Lebesgue-Stieltjes integrable con respecto a F y a la integral $\int_{\mathbb{R}} f d\mu_F$ la llamaremos la integral de Lebesgue-Stieltjes de f con respecto a F. Si $B \in \mathfrak{B}_F(\mathbb{R})$, denotaremos por $\int_B f d\mu_F$ a la integral $\int_{\mathbb{R}} I_B f d\mu_F$.*

Obviamente, la integral de Lebesgue-Stieltjes tiene las propiedades de cualquier integral con respecto a una medida, las cuales fueron expuestas y demostradas en este capítulo 6.

Si $F : [a, b] \to \mathbb{R}$ es una función no decreciente, la vamos a considerar extendida a todo $\mathbb{R}$, definiendo $F(x) = F(a)$ para cualquier $x < a$ y $F(x) = F(b)$ para cualquier $x > b$.

Si $f : [a, b] \to \mathbb{R}$ es una función acotada y $F : [a, b] \to \mathbb{R}$ una función no decreciente, la función f podría integrarse con respecto a F en dos sentidos. Por un lado, f podría ser integrable con respecto a F como integral de Riemann-Stieltjes (Capítulo 3). Por otro lado, f podría ser integrable con respecto a F como integral de Lebesgue (Capítulo 6), para lo cual bastaría que f fuera $\mathfrak{B}_F(\mathbb{R})$-medible ya que es acotada.

Para evitar cualquier confusión, en lo sucesivo, cuando nos refiramos a la integral de Riemann-Stieltjes de una función $f : [a, b] \to \mathbb{R}$ con respecto a una función $g : [a, b] \to \mathbb{R}$, utilizaremos la notación $(RS) \int_a^b f dg$.

Vamos a mostrar que la integral de Lebesgue-Stieltjes es una extensión de la integral de Riemann-Stieltjes. Para ello, probaremos que si f es Riemann-Stieltjes integrable con respecto a F, entonces también es Lebesgue-Stieltjes integrable con respecto a F y que se tiene la igualdad $\int_{[a,b]} f d\mu_F = (RS) \int_a^b f dF$. Además, mostraremos que la integral de Lebesgue-Stieltjes extiende la familia de funciones integrables, para lo cual daremos ejemplos de funciones que son Lebesgue-Stieltjes integrables pero no Riemann-Stieltjes integrables. Como caso particular, tomando $F(x) = x$ para cualquier $x \in [a, b]$, tendremos que la integral de Lebesgue es una extensión de la integral de Riemann.

Teorema 9.1. *Si F es no decreciente y continua y f es Riemann-Stieltjes integrable con respecto a F en el intervalo $[a, b]$, entonces f es medible e integrable y:*

$\int_{[a,b]} f d\mu_F = (RS) \int_a^b f dF.$

Demostración

Como f y F no tienen discontinuidades en común, f es *integrable con respecto a F.

Sea $P_1, P_2, \ldots$ una sucesión de particiones del intervalo $[a, b]$ tales que, para cualquier $n \in \mathbb{N}$, P_{n+1} es un refinamiento de P_n y la norma de P_n tiende a cero cuando n tiende a infinito.

Sea $P_n = \left\{x_0^{(n)}, x_1^{(n)}, \cdots, x_{m_n}^{(n)}\right\}$ y definamos:

$M_i^{(n)} = \sup\left\{f(x) : x \in \left(x_{i-1}^{(n)}, x_i^{(n)}\right]\right\}$, para $i \in \{1, 2, \ldots, m_n\}$,

$m_i^{(n)} = \text{ínf}\left\{f(x) : x \in \left(x_{i-1}^{(n)}, x_i^{(n)}\right]\right\}$, para $i \in \{1, 2, \ldots, m_n\}$,

$\alpha^{(n)} = f(a) I_{\{a\}} + \sum_{i=1}^{m_n} M_i^{(n)} I_{\left(x_{i-1}^{(n)}, x_i^{(n)}\right]}$,

$\beta^{(n)} = f(a) I_{\{a\}} + \sum_{i=1}^{m_n} m_i^{(n)} I_{\left(x_{i-1}^{(n)}, x_i^{(n)}\right]}$.

Como f es *integrable con respecto a F, se tiene:

$\text{lím}_{n \rightsquigarrow \infty} \sum_{i=1}^{m_n} M_i^{(n)} \left[F\left(x_i^{(n)}\right) - F\left(x_{i-1}^{(n)}\right)\right] = \text{lím}_{n \to \infty} \sum_{i=1}^{m_n} m_i^{(n)} \left[F\left(x_i^{(n)}\right) - F\left(x_{i-1}^{(n)}\right)\right]$

$= (RS) \int_a^b f dF.$

Obviamente, $\alpha^{(n)}$ y $\beta^{(n)}$ son funciones medibles para cualquier $n \in \mathbb{N}$ y están acotadas por la misma constante ya que f es acotada. Además, la sucesión de funciones $\left(\alpha^{(n)}\right)_{n \in \mathbb{N}}$ es decreciente y la sucesión $\left(\beta^{(n)}\right)_{n\in\mathbb{N}}$ es creciente; así que $\alpha = \text{lím}_{n\to\infty}\alpha^{(n)}$ y $\beta = \text{lím}_{n\to\infty}\beta^{(n)}$ son también medibles y acotadas. Además:

$\int_{[a,b]}\alpha^{(n)}d\mu_F = f(a)\,\mu_F(\{a\}) + \sum_{i=1}^{m_n} M_i^{(n)}\left[F\left(x_i^{(n)}\right) - F\left(x_{i-1}^{(n)}\right)\right]$

$= \sum_{i=1}^{m_n} M_i^{(n)}\left[F\left(x_i^{(n)}\right) - F\left(x_{i-1}^{(n)}\right)\right],$

$\int_{[a,b]}\beta^{(n)}d\mu_F = f(a)\,\mu_F(\{a\}) + \sum_{i=1}^{m_n} m_i^{(n)}\left[F\left(x_i^{(n)}\right) - F\left(x_{i-1}^{(n)}\right)\right]$

$= \sum_{i=1}^{m_n} m_i^{(n)}\left[F\left(x_i^{(n)}\right) - F\left(x_{i-1}^{(n)}\right)\right],$

$\beta^{(n)} \leq f \leq \alpha^{(n)}$ para cualquier $n\in\mathbb{N}$, de lo cual se sigue que $\beta \leq f \leq \alpha$.

Por el teorema de la convergencia dominada, se tiene:

$\int_{[a,b]}\alpha d\mu_F = \text{lím}_{n\rightsquigarrow\infty}\int_{[a,b]}\alpha^{(n)}d\mu_F$

$= \text{lím}_{n\rightsquigarrow\infty}\sum_{i=1}^{m_n} M_i^{(n)}\left[F\left(x_i^{(n)}\right) - F\left(x_{i-1}^{(n)}\right)\right] = (RS)\int_a^b f dF$

$\int_{[a,b]}\beta d\mu_F = \text{lím}_{n\rightsquigarrow\infty}\int_{[a,b]}\beta^{(n)}d\mu_F$

$= \text{lím}_{n\rightsquigarrow\infty}\sum_{i=1}^{m_n} M_i^{(n)}\left[F\left(x_i^{(n)}\right) - F\left(x_{i-1}^{(n)}\right)\right] = (RS)\int_a^b f dF.$

Así que, $\int_{[a,b]}(\alpha-\beta)\,d\mu_F = 0$. Por lo tanto:

$\mu_F\{x\in[a,b] : \alpha(x) - \beta(x) > 0\} = 0.$

Por lo tanto, $\beta = f = \alpha$ excepto a lo más en un los puntos de un conjunto de medida μ_F cero. Así que f es medible y, como está acotada y $\mu_F([a,b]) < \infty$, es integrable y se tiene:

$\int_{[a,b]}\beta d\mu_F = \int_{[a,b]} f d\mu_F = \int_{[a,b]}\alpha d\mu_F.$

Se concluye entonces que:

$\int_{[a,b]} f d\mu_F = (RS)\int_a^b f dF.$

■

Corolario 9.1. *Si F es no decreciente y f es Riemann-Stieltjes integrable con respecto a F en el intervalo $[a,b]$, entonces f es medible e integrable y:*

$\int_{[a,b]} f d\mu_F = (RS)\int_a^b f dF.$

Demostración

Como f es Riemann-Stieltjes integrable con respecto a F, no tienen discontinuidades en común del mismo lado.

Expresemos F como la suma $F^d + F^i + F^c$, donde F^d es una función no decreciente, continua por la derecha, que crece únicamente mediante saltos y tal que $F^d(x) - F^d(x-) = F(x) - F(x-)$ para cualquier $x \in \mathbb{R}$, F^i es una función no decreciente, continua por la izquierda, que crece únicamente mediante saltos y tal que $F^i(x+) - F^i(x) = F(x+) - F(x)$ para cualquier $x \in \mathbb{R}$, y F^c es una función no decreciente y continua.

Obsérvese que $\mu_F = \mu_{F^d} + \mu_{F^i} + \mu_{F^c}$ y que, si $D = \{x_1, x_2, \ldots\}$ es el conjunto donde F es discontinua, entonces $\mu_F(A) > 0$ para cualquier subconjunto $A \subset D$ y $\mu_{F^d}(D^c) = \mu_{F^i}(D^c) = 0$, así que la familia de conjuntos $B \in \mathfrak{B}(\mathbb{R})$ tales que $B \subset D^c$ y $\mu_F(B) = 0$, coincide con la familia de conjuntos $B \in \mathfrak{B}(\mathbb{R})$ tales que $B \subset D^c$ y $\mu_{F^c}(B) = 0$. Por lo tanto, la σ-álgebra completada $\mathfrak{B}_F(\mathbb{R})$ coincide con la σ-álgebra completada $\mathfrak{B}_{F^c}(\mathbb{R})$.

Como F^d es una función no decreciente, continua por la derecha y crece únicamente mediante saltos y f es una función acotada y continua por la izquierda en los puntos donde F^d es discontinua, entonces, por la proposición 4.6, f es integrable con respecto a F^d y se tiene $(RS)\int_a^b f dF^d = \sum_{\{x \in D\}} f(x)\left[F^d(x) - F^d(x-)\right]$.

De la misma manera, como F^i es una función no decreciente, continua por la izquierda y crece únicamente mediante saltos y f es una función acotada y continua por la derecha en los puntos donde F^d es discontinua, entonces, por la proposición 4.7, f es integrable con respecto a F^d y se tiene $(RS)\int_a^b f dF^i = \sum_{\{x \in D\}} f(x)\left[F^i(x+) - F^i(x)\right]$.

Además, como f es integrable con respecto a F^d, con respecto a F^i y con respecto a F, también es integrable con respecto a F^c. Así que, por la proposición anterior, f es $\mathfrak{B}_F(\mathbb{R})$-medible y μ_{F^c}-integrable, y se tiene:

$$\int_{[a,b]} f d\mu_{F^c} = (RS)\int_a^b f dF^c.$$

Además, como f está acotada, es μ_{F^d}-integrableμ_{F^i}-integrable y y se tiene:

$$\int_{[a,b]} f d\mu_{F^d} = \sum_{\{x \in D\}} f(x)\mu_{F^d}(\{x\}) = \sum_{\{x \in D\}} f(x)\left[F^d(x) - F^d(x-)\right],$$

$$\int_{[a,b]} f d\mu_{F^i} = \sum_{\{x \in D\}} f(x)\mu_{F^i}(\{x\}) = \sum_{\{x \in D\}} f(x)\left[F^i(x+) - F^i(x)\right].$$

Por lo tanto:

$$\int_{[a,b]} f d\mu_F = \int_{[a,b]} f d\mu_{F^d} + \int_{[a,b]} f d\mu_{F^i} + \int_{[a,b]} f d\mu_{F^c}$$

$$= (RS)\int_a^b f dF^d + (RS)\int_a^b f dF^i + (RS)\int_a^b f dF^c = (RS)\int_a^b f dF.$$

■

9.1. Propiedades de la integral de Lebesgue Stieltjes

Ejemplos de funciones Lebesgue-Stieltjes integrables que no son Riemann-Stieltjes integrables, con respecto a una función no decreciente, hay muchos. Por ejemplo, toda función de variación acotada $f : (\mathbb{R}, \mathfrak{B}([a, b])) \to \mathbb{R}$ es medible ya que se puede expresar como la diferencia de dos funciones no decrecientes, así que, cualquiera que sea la función no decreciente F, f es Lebesgue-Stieltjes integrable con respecto a F. En cambio, para que f sea Riemann-Stieltjes integrable con respecto a F se requiere que f y F no tengan discontinuidades en común del mismo lado. Otro ejemplo se obtiene si definimos $f = I_B$, donde $B \in \mathfrak{B}([a, b])$. Una función de este tipo es Lebesgue-Stieltjes integrable con respecto a cualquier función F no decreciente y su integral está dada por $\mu_F(B)$. Por otra parte, si B es el conjunto de números racionales en el intervalo $[a, b]$, entonces I_B no es Riemann-Stieltjes integrable con respecto a cualquier función no constante F no decreciente y continua.

Por otra parte, si la función F no es de variación acotada, no genera una medida μ_F sobre $(\mathbb{R}, \mathfrak{B}([a, b]))$, así que la integral de Lebesgue-Stieltjes, con respecto a F, de cualquier función acotada $f : [a, b] \to \mathbb{R}$ no está definida, pero f podría ser Riemann-Stieltjes integrable con respecto a F. Tal es el caso si F es una función continua que no es de variación acotada y f una función de variación acotada.

Restringiéndonos a funciones F que son de variación acotada, se tienen algunos resultados interesantes. Uno es una extensión de uno de los ejemplos anteriores: Toda función de variación acotada $f : [a, b] \to \mathbb{R}$ es Lebesgue-Stieltjes integrable con respecto a cualquier función de variación acotada $F : [a, b] \to \mathbb{R}$.

Por otra parte, algunos de los resultados que obtuvimos para la integral de Riemann-Stieltjes podemos ahora formularlos de una manera más general.

Nos vamos a restringir a considerar la integral de Lebesgue-Stieltjes de una función f con respecto a una función no decreciente F para el caso de funciones no decrecientes y continuas por la derecha definidas sobre el intervalo $[0, \infty)$, el cual, como ya lo mencionamos con anterioridad, también lo denotamos por $\mathbb{R}^+$. Sin embargo, para evitar nuevos enunciados, si $F : [0, \infty) \to \mathbb{R}$ es una función no decreciente, la vamos a considerar extendida a todo $\mathbb{R}$, definiendo $F(x) = F(0)$ para cualquier $x < 0$. Así que la medida μ_F generada por F la consideraremos definida sobre el espacio $(\mathbb{R}, \mathfrak{B}_F(\mathbb{R}))$ definido al inicio de la sección anterior.

TEOREMA 9.2. *Sea $F : [0, \infty) \to \mathbb{R}$ una función no decreciente continua por la derecha y $f : \mathbb{R} \to \mathbb{R}$ una función integrable con respecto a μ_F sobre cualquier conjunto boreliano acotado, entonces la función $H : [0, \infty) \to \mathbb{R}$ definida por:*

$$H(t) = \int_{[0,t]} f d\mu_F$$

es continua por la derecha.

Además, si F es continua, entonces H es continua.

Demostración

Primero observemos que, como $\mu_F(\{0\}) = 0$, se tiene:

$\int_{[0,t]} f d\mu_F = \int_{(0,t]} f d\mu_F$.

Sea $u \in [0, \infty)$ y $(u_n)_{n\in\mathbb{N}}$ una sucesión decreciente de números reales tal que $\lim_{n\rightsquigarrow\infty} u_n = u$. Entonces:

$I_{(u,u_n]} |f| \leq |f|$ y $\lim_{n\rightsquigarrow\infty} I_{(u,u_n]} |f| = 0$. Así que, por el teorema de la convergencia dominada:

$\lim_{n\rightsquigarrow\infty} |H(u_n) - H(u)| = \lim_{n\rightsquigarrow\infty} \left| \int_{(u,u_n]} f d\mu_F \right|$

$\leq \lim_{n\rightsquigarrow\infty} \int_{(u,u_n]} |f| d\mu_F = \lim_{n\rightsquigarrow\infty} \int_{\mathbb{R}} I_{(u,u_n]} |f| d\mu_F = 0$.

Por lo tanto, H es continua por la derecha en u.

Si F es continua, sea $u \in (0, \infty)$ y $(u_n)_{n\in\mathbb{N}}$ una sucesión creciente de números reales positivos tal que $\lim_{n\rightsquigarrow\infty} u_n = u$. Entonces:

$I_{(u_n,u]} |f| \leq |f|$ y $\lim_{n\rightsquigarrow\infty} I_{(u_n,u]} |f| = I_{\{u\}} |f|$. Así que, por el teorema de la convergencia dominada:

$\lim_{n\rightsquigarrow\infty} |H(u) - H(u_n)| = \lim_{n\rightsquigarrow\infty} \left| \int_{(u_n,u]} f d\mu_F \right|$

$\leq \lim_{n\rightsquigarrow\infty} \int_{(u_n,u]} |f| d\mu_F = \int_{\mathbb{R}} I_{\{u\}} |f| d\mu_F = \mu_F(\{u\}) |f|(u) = 0$.

Por lo tanto, H es continua por la izquierda en u.

∎

Sea $g : [0, \infty) \to \mathbb{R}$ una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto y $f : \mathbb{R} \to \mathbb{R}$ una función acotada sobre los conjuntos acotados. Entonces, si $F_1 : [0, \infty) \to \mathbb{R}$, $F_2 : [0, \infty) \to \mathbb{R}$, $F_1' : [0, \infty) \to \mathbb{R}$ y $F_2' : [0, \infty) \to \mathbb{R}$ son funciones no decrecientes continuas por la derecha tales que $g = F_1 - F_2 = F_1' - F_2'$, definamos:

$G_1 = F_1 + F_2'$,

$G_2 = F_1' + F_2$.

Entonces, para cualquier intervalo $(a, b]$, se tiene:

$\mu_{G_1}((a, b]) = \mu_{G_2}((a, b])$.

Así que, $\mu_{G_1}(B) = \mu_{G_2}(B)$ para cualquier conjunto acotado $B \in \mathfrak{B}(\mathbb{R})$ y, entonces, $\int_B f d\mu_{G_1} = \int_B f d\mu_{G_2}$.

Por lo tanto:

$\int_B f d\mu_{F_1} - \int_B f d\mu_{F_2} = \int_B f d\mu_{F_1'} - \int_B f d\mu_{F_2'}$.

La situación es distinta cuando f no es acotada ya que si $F : [0, \infty) \to \mathbb{R}$ es cualquier función no decreciente continua por la derecha, $G_1 = F_1 + F$ y $G_2 = F_2 + F$ entonces G_1 y G_2 son funciones no decrecientes continuas por la derecha y $g = G_1 - G_2$. Así que, para tener la igualdad $\int_B f d\mu_{F_1} - \int_B f d\mu_{F_2} = \int_B f d\mu_{G_1} - \int_B f d\mu_{G_2}$, por lo menos f tendría que ser integrable, sobre B, con respecto a μ_F.

Consideremos, como ejemplo, el caso siguiente:

Sean $g : [0, \infty) \to \mathbb{R}$, $F_1 : [0, \infty) \to \mathbb{R}$ y $F_2 : [0, \infty) \to \mathbb{R}$ definidas de la siguiente manera:

$$g(s) = \begin{cases} 0 & \text{si } s \in (-\infty, 1) \\ \frac{s-1}{s^2} & \text{si } s \in [1, \infty) \end{cases}$$

$$F_1(s) = \begin{cases} 0 & \text{si } s \in (-\infty, 1) \\ 1 - \frac{1}{s^2} & \text{si } s \in [1, \infty) \end{cases}$$

$$F_2(s) = \begin{cases} 0 & \text{si } s \in (-\infty, 1) \\ 1 - \frac{1}{s} & \text{si } s \in [1, \infty) \end{cases}$$

Entonces $g : [0, \infty) \to \mathbb{R}$ es una función continua, de variación acotada sobre cualquier intervalo compacto, F_1 y F_2 son funciones no decrecientes continuas y $g = F_1 - F_2$.

Definamos $f : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:

$$f(s) = \begin{cases} 0 & \text{si } s \in (-\infty, 1) \\ 2^{\frac{k}{2}} & \text{si } s \in \left(n - \frac{1}{2^{k-1}}, n - \frac{1}{2^k}\right], \text{ donde } k \in \mathbb{N} \text{ y } n \in \{2, 3, \ldots\} \\ 0 & \text{si } s \in \mathbb{N} \end{cases}$$

f es entonces una función escalonada la cual tiende a infinito, cuando k tiende a infinito, en cada intervalo $(n, n+1)$, donde $n \in \mathbb{N}$.

Para cualquier $n \in \mathbb{N}$, se tiene:

$\int_n^{n+1} f(s)\, ds = \sum_{k=1}^{\infty} 2^{\frac{k}{2}} \left(\frac{1}{2^{k-1}} - \frac{1}{2^k}\right) = \sum_{k=1}^{\infty} 2^{\frac{k}{2}} \frac{1}{2^k}$

$= \sum_{k=1}^{\infty} \frac{1}{2^{\frac{k}{2}}} = \frac{\frac{1}{\sqrt{2}}}{1 - \frac{1}{\sqrt{2}}} = \frac{1}{\sqrt{2}-1}$,

$\int_n^{n+1} f d\mu_{F_1} = \int_n^{n+1} \frac{2}{s^3} f(s)\, ds \leq 2 \int_n^{n+1} f(s)\, ds = \frac{2}{\sqrt{2}-1}$,

$\int_n^{n+1} f d\mu_{F_2} = \int_n^{n+1} \frac{1}{s^2} f(s)\, ds \leq \int_n^{n+1} f(s)\, ds = \frac{1}{\sqrt{2}-1}$.

Si $F : [0, \infty) \to \mathbb{R}$ es cualquier función no decreciente continua por la derecha, entonces:

$g = (F_1 + F) - (F_2 + F)$.

Definamos $F : [0, \infty) \to \mathbb{R}$ de tal forma que:

1. $\mu_F\left(\left(n - \frac{1}{2^{k-1}}, n - \frac{1}{2^k}\right]\right) = \frac{1}{\left(2^{\frac{1}{3}}\right)^k}$

para cualesquiera $k \in \mathbb{N}$ y $n \in \{2, 3, \ldots\}$.

2. $\mu_F((-\infty, 1)) = 0$ y $\mu_F(\{n\}) = 0$ para cualquier $n \in \mathbb{N}$.

3. La medida total de cada intervalo $\left(n - \frac{1}{2^{k-1}}, n - \frac{1}{2^k}\right]$ queda repartida uniformemente en ese intervalo.

Se tiene, para cualquier $n \in \mathbb{N}$:

$\mu_F((n, n+1]) = \sum_{k=1}^{\infty} \frac{1}{\left(2^{\frac{1}{3}}\right)^k} = \frac{\frac{1}{2^{\frac{1}{3}}}}{1 - \frac{1}{2^{\frac{1}{3}}}} = \frac{1}{2^{\frac{1}{3}} - 1} = \frac{1}{\sqrt[3]{2} - 1}$.

Así que:

$F(s) = 0$ para cualquier $s \in [0, 1)$.

$F(s) = \frac{n-1}{\sqrt[3]{2}-1} + \sum_{j=1}^{k-1} \frac{1}{2^{\frac{j}{3}}} + \frac{1}{\left(2^{\frac{1}{3}}\right)^k} \frac{s - \left(n - \frac{1}{2^{k-1}}\right)}{\frac{1}{2^k}} = \frac{1}{\sqrt[3]{2}-1} 2^{-\frac{k-1}{3}} \left(2^{\frac{k-1}{3}} - 1\right) + \left(2^{\frac{2}{3}}\right)^k \left[s - \left(n - \frac{1}{2^{k-1}}\right)\right]$

para cualesquiera $k \in \mathbb{N}$, $n \in \{2, 3, \ldots\}$ y $s \in \left(n - \frac{1}{2^{k-1}}, n - \frac{1}{2^k}\right]$.

$F(n) = \frac{n-1}{\sqrt[3]{2}-1}$ para cualquier $n \in \mathbb{N}$.

Para cualquier $n \in \mathbb{N}$, se tiene:

$\int_n^{n+1} f d\mu_F = \sum_{k=1}^{\infty} \frac{2^{\frac{k}{2}}}{2^{\frac{k}{3}}} = \sum_{k=1}^{\infty} 2^{\frac{k}{6}} = \infty.$

Así que:

$\int_n^{n+1} f d\mu_{(F_1+F)} = \int_n^{n+1} f d\mu_{(F_2+F)} = \infty.$

Definición 9.2. *Sea $g : [0, \infty) \to \mathbb{R}$ una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto. Si $f : \mathbb{R} \to \mathbb{R}$ es una función acotada sobre los conjuntos acotados y $B \in \mathfrak{B}(\mathbb{R})$ es un conjunto acotado, definimos la integral de f con respecto a g, sobre el conjunto B, de la siguiente manera:*

$\int_B f dg = \int_B f d\mu_{F_1} - \int_B f d\mu_{F_2},$

donde $F_1 : [0, \infty) \to \mathbb{R}$ y $F_2 : [0, \infty) \to \mathbb{R}$ es cualquier par de funciones no decrecientes continuas por la derecha tales que $g = F_1 - F_2$.

Si a y b son dos números reales tales que $a < b$, también denotaremos por $\int_a^b f dg$ a la integral $\int_{(a,b]} f dg$. De igual forma, si $F : [0, \infty) \to \mathbb{R}$ es una función no decreciente continua por la derecha, también denotaremos por $\int_a^b f d\mu_F$, o por $\int_a^b f dF$, a la integral $\int_{(a,b]} f d\mu_F$ y por $\int_B f dF$ a la integral $\int_B f d\mu_F$.

La integral $\int_B fdg$ está bien definida ya que, como lo mencionamos antes, si $F_1^{'} : [0, \infty) \to \mathbb{R}$ y $F_2^{'} : [0, \infty) \to \mathbb{R}$ es otro par de funciones no decrecientes continuas por la derecha tales que $g = F_1^{'} - F_2^{'}$ entonces:

$\int_B fd\mu_{F_1} - \int_B fd\mu_{F_2} = \int_B fd\mu_{F_1^{'}} - \int_B fd\mu_{F_2^{'}}$.

Por otra parte, la notación $\int_B fdF$ para la integral $\int_B fd\mu_F$ es consistente con la definición anterior ya que si g es no decreciente, entonces, de acuerdo con la definición 9.2, se tiene:

$\int_B fdg = \int_B fd\mu_g$.

El siguiente resultado es inmediato:

Proposición 9.1. *Sea $g : [0, \infty) \to \mathbb{R}$ una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto, $f : \mathbb{R} \to \mathbb{R}$ y $h : \mathbb{R} \to \mathbb{R}$ dos funciones acotadas sobre los conjuntos acotados, $a, b \in \mathbb{R}$ y $B \in \mathfrak{B}(\mathbb{R})$ un conjunto acotado, entonces:*

$\int_B (af + bh)\, dg = a \int_B fdg + b \int_B hdg.$

Proposición 9.2. *Sea $g : [0, \infty) \to \mathbb{R}$ una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto, y $h : \mathbb{R} \to \mathbb{R}$ una función acotada sobre los conjuntos acotados. Si $(f_n)_{n\in\mathbb{N}}$ es una sucesión de funciones medibles tales que $|f_n| \leq |h|$ y f es una función medible tal que $f = \text{lím}_{n\rightsquigarrow\infty} f_n$ excepto a lo más en un conjunto de medida cero, entonces f y f_n, para cualquier $n \in \mathbb{N}$, son funciones acotadas sobre los conjuntos acotados, y se tiene:*

$\int_B fdg = \text{lím}_{n\rightsquigarrow\infty} \int_B f_n dg$

para cualquier conjunto acotado $B \in \mathfrak{B}(\mathbb{R})$.

Demostración

Sean $F_1 : [0, \infty) \to \mathbb{R}$ y $F_2 : [0, \infty) \to \mathbb{R}$ dos funciones no decrecientes continuas por la derecha, tales que $g = F_1 - F_2$.

Se tiene $|f| \leq |h|$, así que, f y f_n, para cualquier $n \in \mathbb{N}$, y cualquier conjunto acotado $B \in \mathfrak{B}(\mathbb{R})$, se tiene:

Así que f y f_n, para cualquier $n \in \mathbb{N}$, son funciones acotadas sobre los conjuntos acotados.

Además, por el teorema de la convergencia dominada, se tiene:

$\int_B fdg = \int_B fd\mu_{F_1} - \int_B fd\mu_{F_2} = \text{lím}_{n\rightsquigarrow\infty} \int_B f_n d\mu_{F_1} - \text{lím}_{n\rightsquigarrow\infty} \int_B f_n d\mu_{F_2}$

$= \text{lím}_{n\rightsquigarrow\infty} \left(\int_B f_n d\mu_{F_1} - \int_B f_n d\mu_{F_2}\right) = \text{lím}_{n\to\infty} \int_B f_n dg.$

∎

TEOREMA 9.3. *Sea $g : [0, \infty) \to \mathbb{R}$ una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto, y $f : \mathbb{R} \to \mathbb{R}$ una función acotada sobre los conjuntos acotados, entonces la función $G : [0, \infty) \to \mathbb{R}$ definida por:*

$G(t) = \int_{[0,t]} f dg$

es continua por la derecha y de variación acotada sobre cualquier intervalo compacto.

Además, si g es continua, entonces G es continua.

Demostración

Sean $F_1 : [0, \infty) \to \mathbb{R}$ y $F_2 : [0, \infty) \to \mathbb{R}$ dos funciones no decrecientes continuas por la derecha, tales que $g = F_1 - F_2$.

Sea $t \in (0, \infty)$ y $P = \{x_0, x_1, \dots, x_n\}$ una partición del intervalo $[0, t]$. Entonces:

$\sum_{k=1}^{n} |G(x_k) - G(x_{k-1})| = \sum_{k=1}^{n} \left| \int_{(x_{k-1},x_k]} f d\mu_{F_1} - \int_{(x_{k-1},x_k]} f d\mu_{F_2} \right|$

$\leq \sum_{k=1}^{n} \int_{(x_{k-1},x_k]} |f| \, d\mu_{F_1} + \sum_{k=1}^{n} \int_{(x_{k-1},x_k]} |f| \, d\mu_{F_2} = \int_{(0,t]} |f| \, d\mu_{F_1} + \int_{(0,t]} |f| \, d\mu_{F_2}.$

Así que:

$V_G[a, b] = \sup \{V_G(P) : P \text{ es una partición de } [a, b]\} \leq \int_{(0,t]} |f| \, d\mu_{F_1} + \int_{(0,t]} |f| \, d\mu_{F_2} < \infty.$

Así que G es de variación acotada sobre cualquier intervalo compacto.

Por otra parte, se tiene:

$\int_{[0,t]} f dg = \int_{[0,t]} f d\mu_{F_1} - \int_{[0,t]} f d\mu_{F_2}.$

Además, si g es continua, podemos tomar F_1 y F_2 continuas.

Así que, por el teorema 9.2, G es continua por la derecha y, si g es continua, G también lo es.

■

En el capítulo 3 estudiamos la parte continua y la parte de saltos de una función no decreciente, así como de una función de variación acotada, definida sobre $\mathbb{R}$. Ahora estamos trabajando con ese mismo tipo de funciones, pero definidas sobre $\mathbb{R}^+$. En este caso la definición de la parte continua y de la parte de saltos de la función es más simple, así que las vamos a definir directamente.

Si $F : [0, \infty) \to \mathbb{R}$ es una función no decreciente continua por la derecha, la función $F^d : [0, \infty) \to \mathbb{R}$ definida por:

$F^d(t) = \sum_{s \in [0,t]} [F(s) - F(s-)],$

entonces F es no decreciente y continua por la derecha.

Además, la función $F^c = F - F^d$ es no decreciente y continua.

Así que, si $g : [0, \infty) \to \mathbb{R}$ es una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto, entonces la fiunción $g^d : [0, \infty) \to \mathbb{R}$ definida por:

$g^d(t) = \sum_{s\in[0,t]} [g(s) - g(s-)]$

es continua por la derecha y de variación acotada sobre cualquier intervalo compacto y la función $g^c = g - g^d$ es continua y de variación acotada sobre cualquier intervalo compacto.

Además, si $F_1 : [0, \infty) \to \mathbb{R}$ y $F_2 : [0, \infty) \to \mathbb{R}$ son dos funciones no decrecientes continuas por la derecha tales que $g = F_1 - F_2$, entonces:

$g^d = F_1^d - F_2^d,$

$g^c = F_1^c - F_2^c.$

Definición 9.3. *Si $h : [0, \infty) \to \mathbb{R}$ es una función continua por la derecha, no decreciente o de variación acotada sobre cualquier intervalo compacto, las funciones h^d y h^c serán llamadas la parte discreta y la parte continua, respectivamente, de h.*

Proposición 9.3. *Sea $g : [0, \infty) \to \mathbb{R}$ una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto, y $f : \mathbb{R} \to \mathbb{R}$ una función acotada sobre los conjuntos acotados, entonces la serie:*

$\sum_{s\in[0,t]} f(s) [g(s) - g(s-)]$

es absolutamente convergente para cualquier $t \in [0, \infty)$.

Demostración

Sean $F_1 : [0, \infty) \to \mathbb{R}$ y $F_2 : [0, \infty) \to \mathbb{R}$ dos funciones no decrecientes continuas por la derecha, tales que $g = F_1 - F_2$.

Como f es integrable con respecto a $\mu_{F_1^d}$ y con respecto a $\mu_{F_2^d}$ sobre los conjuntos borelianos acotados, las series:

$\sum_{s\in[0,t]} |f(s)| [F_1(s) - F_1(s-)] = \int_0^t |f| \, d\mu_{F_1^d},$

$\sum_{s\in[0,t]} |f(s)| [F_2(s) - F_2(s-)] = \int_0^t |f| \, d\mu_{F_2^d},$

son convergentes para cualquier $t \in [0, \infty)$.

Además:

$\sum_{s\in[0,t]} |f(s)| |g(s) - g(s-)| = \sum_{s\in[0,t]} |f(s)| |F_1(s) - F_1(s-) - [F_2(s) - F_2(s-)]|$

$\leq \sum_{s\in[0,t]} |f(s)| |F_1(s) - F_1(s-)| + \sum_{s\in[0,t]} |f(s)| |F_2(s) - F_2(s-)| < \infty$

para cualquier $t \in [0,\infty)$.

■

TEOREMA 9.4. *Sea $g : [0,\infty) \to \mathbb{R}$ una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto, y $f : \mathbb{R} \to \mathbb{R}$ una función acotada sobre los conjuntos acotados, entonces:*

$\int_0^t f dg = \int_0^t f dg^c + \sum_{s\in[0,t]} f(s) [g(s) - g(s-)]$

para cualquier $t \in [0,\infty)$.

Demostración

Sean $F_1 : [0,\infty) \to \mathbb{R}$ y $F_2 : [0,\infty) \to \mathbb{R}$ dos funciones no decrecientes continuas por la derecha, tales que $g = F_1 - F_2$.

Entonces:

$\int_0^t f dg = \int_0^t f dF_1 - \int_0^t f dF_2 = \int_0^t f dF_1^c + \int_0^t f dF_1^d - \left(\int_0^t f dF_2^c + \int_0^t f dF_2^d\right)$

$= \int_0^t f dF_1^c - \int_0^t f dF_2^c + \int_0^t f dF_1^d - \int_0^t f dF_2^d$

$= \int_0^t f dg^c + \int_0^t f dF_1^d - \int_0^t f dF_2^d$

$= \int_0^t f dg^c + \sum_{s\in[0,t]} f(s) [F_1(s) - F_1(s-)] - \sum_{s\in[0,t]} f(s) [F_2(s) - F_2(s-)]$

$= \int_0^t f dg^c + \sum_{s\in[0,t]} f(s) \{F_1(s) - F_2(s) - [F_1(s-) - F_2(s-)]\}$

$= \int_0^t f dg^c + \sum_{s\in[0,t]} f(s) [g(s) - g(s-)]$.

■

Sean $g : [0,\infty) \to \mathbb{R}$ una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto, $F_1 : [0,\infty) \to \mathbb{R}$ y $F_2 : [0,\infty) \to \mathbb{R}$ dos funciones no decrecientes continuas por la derecha, tales que $g = F_1 - F_2$, $f : \mathbb{R} \to \mathbb{R}$ una función acotada sobre los conjuntos acotados, y $G : [0,\infty) \to \mathbb{R}$ definida por $G(t) = \int_0^t f dg$. Se tiene entonces:

$G(t) = \int_0^t f dg = \int_0^t f d\mu_{F_1} - \int_0^t f d\mu_{F_2} = \int_0^t f^+ d\mu_{F_1} - \int_0^t f^- d\mu_{F_1} - \left(\int_0^t f^+ d\mu_{F_2} - \int_0^t f^- d\mu_{F_2}\right)$

$= \int_0^t f^+ d\mu_{F_1} + \int_0^t f^- d\mu_{F_2} - \left(\int_0^t f^- d\mu_{F_1} + \int_0^t f^+ d\mu_{F_2}\right)$.

Definamos $G_1 [0,\infty) \to \mathbb{R}$ y $G_2 : [0,\infty) \to \mathbb{R}$ de la siguiente manera:

$G_1(t) = \int_0^t f^+ d\mu_{F_1} + \int_0^t f^- d\mu_{F_2}$,

$G_2(t) = \int_0^t f^- d\mu_{F_1} + \int_0^t f^+ d\mu_{F_2}$.

Entonces G_1 y G_2 son funciones no decrecientes continuas por la derecha y $G = G_1 - G_2$.

Así que, si $B \in \mathfrak{B}(\mathbb{R})$ es un conjunto acotado y $h : \mathbb{R} \to \mathbb{R}$ es una función acotada sobre los conjuntos acotados, se tiene:

$\int_B hdG = \int_B hd\mu_{G_1} - \int_B hd\mu_{G_2}$.

TEOREMA 9.5. *Sea $g : [0, \infty) \to \mathbb{R}$ una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto, $f : \mathbb{R} \to \mathbb{R}$ una función acotada sobre los conjuntos acotados y $G : [0, \infty) \to \mathbb{R}$ definida por $G(t) = \int_0^t f dg$. Si $h : \mathbb{R} \to \mathbb{R}$ es una función acotada sobre los conjuntos acotados, entonces:*

$\int_B hdG = \int_B hf dg$

para cualquier conjunto acotado $B \in \mathfrak{B}(\mathbb{R})$.

Demostración

Sean $F_1 : [0, \infty) \to \mathbb{R}$ y $F_2 : [0, \infty) \to \mathbb{R}$ dos funciones no decrecientes continuas por la derecha, tales que $g = F_1 - F_2$.

Definamos $G_1 : [0, \infty) \to \mathbb{R}$, $G_2 : [0, \infty) \to \mathbb{R}$ y las medidas ν_1^+, ν_2^+, ν_1^- y ν_2^-, sobre $\mathfrak{B}(\mathbb{R})$, de la siguiente manera:

$G_1(t) = \int_0^t f^+ d\mu_{F_1} + \int_0^t f^- d\mu_{F_2}$,

$G_2(t) = \int_0^t f^- d\mu_{F_1} + \int_0^t f^+ d\mu_{F_2}$,

$\nu_1^+(B) = \int_B f^+ d\mu_{F_1}$,

$\nu_2^+(B) = \int_B f^+ d\mu_{F_2}$,

$\nu_1^-(B) = \int_B f^- d\mu_{F_1}$,

$\nu_2^-(B) = \int_B f^- d\mu_{F_2}$.

Entonces G_1 y G_2 son funciones no decrecientes continuas por la derecha y $G = G_1 - G_2$.

Además:

$\int_B hdG = \int_B hd\mu_{G_1} - \int_B hd\mu_{G_2} = \int_B hd\nu_1^+ + \int_B hd\nu_2^- - \int_B hd\nu_1^- - \int_B hd\nu_2^+$

$= \int_B h^+ d\nu_1^+ - \int_B h^- d\nu_1^+ + \int_B h^+ d\nu_2^- - \int_B h^- d\nu_2^- - \int_B h^+ d\nu_1^-$

$+ \int_B h^- d\nu_1^- - \int_B h^+ d\nu_2^+ + \int_B h^- d\nu_2^+$

$= \int_B h^+ f^+ d\mu_{F_1} - \int_B h^- f^+ d\mu_{F_1} + \int_B h^+ f^- d\mu_{F_2} - \int_B h^- f^- d\mu_{F_2} - \int_B h^+ f^- d\mu_{F_1}$

$+\int_B h^- f^- d\mu_{F_1} - \int_B h^+ f^+ d\mu_{F_2} + \int_B h^- f^+ d\mu_{F_2}$

$= \int_B h^+ f^+ d\mu_{F_1} - \int_B h^+ f^- d\mu_{F_1} - \int_B h^- f^+ d\mu_{F_1} + \int_B h^- f^- d\mu_{F_1} - \int_B h^+ f^+ d\mu_{F_2}$

$+\int_B h^+ f^- d\mu_{F_2} + \int_B h^- f^+ d\mu_{F_2} - \int_B h^- f^- d\mu_{F_2}$

$= \int_B h^+ f d\mu_{F_1} - \int_B h^- f d\mu_{F_1} - \int_B h^+ f d\mu_{F_2} + \int_B h^- f d\mu_{F_2}$

$= \int_B hf d\mu_{F_1} - \int_B hf d\mu_{F_2} = \int_B hf dg.$

■

9.2. Fórmula de integración por partes

TEOREMA 9.6. *Sean $F : [0,\infty) \to \mathbb{R}$ y $G : [0,\infty) \to \mathbb{R}$ dos funciones no decrecientes continuas por la derecha, entonces:*

$$F(t)\,G(t) = F(0)\,G(0) + \int_0^t FdG + \int_0^t GdF - \sum_{s\in[0,t]} [F(s) - F(s-)]\,[G(s) - G(s-)]$$

para cualquier $t \in [0,\infty)$.

Demostración

Sea μ la medida producto $\mu_F \times \mu_G$, definida sobre $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$, $t \in (0,\infty)$ y $C_t = [0,t]\times[0,t]$. Entonces, aplicando el teorema de Fubini, se tiene:

$[F(t) - F(0)]\,[G(t) - G(0)] = \mu(C_t)$

$= \mu(\{(u,v) \in \mathbb{R}^2 : 0 \le u < v \le t\}) + \mu(\{(u,v) \in \mathbb{R}^2 : 0 \le v \le u \le t\})$

$= \int_{[0,t]} \left(\int_{[0,v)} dF(u)\right) dG(v) + \int_{[0,t]} \left(\int_{[0,u]} dG(v)\right) dF(u)$

$= \int_{[0,t]} [F(v-) - F(0)]\,dG(v) + \int_{[0,t]} [G(u) - G(0)]\,dF(u)$

$= \int_{[0,t]} F(v-)\,dG(v) - F(0)\,[G(t) - G(0)] + \int_{[0,t]} G(u)\,dF(u) - G(0)\,[F(t) - F(0)].$

Así que:

$F(t)\,G(t) = F(0)\,G(0) + \int_0^t F(v-)\,dG(v) + \int_0^t G(u)\,dF(u)$

$= F(0)\,G(0) + \int_0^t F(v)\,dG(v) + \int_0^t G(u)\,dF(u) - \int_0^t [F(v) - F(v-)]\,dG(v)$

$= F(0)\,G(0) + \int_0^t FdG + \int_0^t GdF - \sum_{v\in[0,t]} [F(v) - F(v-)]\,[G(v) - G(v-)].$

■

TEOREMA 9.7. *Sean $g : [0,\infty) \to \mathbb{R}$ y $h : [0,\infty) \to \mathbb{R}$ dos funciones continua por la derecha, de variación acotada sobre cualquier intervalo compacto, entonces:*

$g(t)\,h(t) = g(0)\,h(0) + \int_0^t gdh + \int_0^t hdg - \sum_{s\in[0,t]} [g(s) - g(s-)]\,[h(s) - h(s-)]$

para cualquier $t \in [0, \infty)$.

Demostración

Sean $F_1 : [0,\infty) \to \mathbb{R}$, $F_2 : [0,\infty) \to \mathbb{R}$, $G_1 : [0,\infty) \to \mathbb{R}$ y $G_2 : [0,\infty) \to \mathbb{R}$ funciones no decrecientes continuas por la derecha, tales que $g = F_1 - F_2$ y $h = G_1 - G_2$. Entonces, para cualesquiera $s, t \in [0,\infty)$, se tiene:

$\int_0^t gdh = \int_0^t gdG_1 - \int_0^t gdG_2 = \int_0^t F_1dG_1 - \int_0^t F_2dG_1 - \int_0^t F_1dG_2 + \int_0^t F_2dG_2,$

$\int_0^t hdg = \int_0^t hdF_1 - \int_0^t hdF_2 = \int_0^t G_1dF_1 - \int_0^t G_2dF_1 - \int_0^t G_1dF_2 + \int_0^t G_2dF_2,$

$[g(s) - g(s-)]\,[h(s) - h(s-)]$

$= [F_1(s) - F_2(s) - F_1(s-) + F_2(s-)]\,[G_1(s) - G_2(s) - G_1(s-) + G_2(s-)]$

$= [F_1(s) - F_1(s-)]\,[G_1(s) - G_1(s-)] + [F_2(s) - F_2(s-)]\,[G_2(s) - G_2(s-)]$

$- [F_1(s) - F_1(s-)]\,[G_2(s) - G_2(s-)] - [F_2(s) - F_2(s-)]\,[G_1(s) - G_1(s-)]$.

Así que:

$g(t)\,h(t) = [F_1(t) - F_2(t)]\,[G_1(t) - G_2(t)]$

$= F_1(t)\,G_1(t) + F_2(t)\,G_2(t) - F_1(t)\,G_2(t) - F_2(t)\,G_1(t)$

$= F_1(0)\,G_1(0) + \int_0^t F_1dG_1 + \int_0^t G_1dF_1 - \sum_{s\in[0,t]} [F_1(s) - F_1(s-)]\,[G_1(s) - G_1(s-)]$

$+F_2(0)\,G_1(0) + \int_0^t F_2dG_2 + \int_0^t G_2dF_2 - \sum_{s\in[0,t]} [F_2(s) - F_2(s-)]\,[G_2(s) - G_2(s-)]$

$-F_1(0)\,G_2(0) - \int_0^t F_1dG_2 - \int_0^t G_2dF_1 + \sum_{s\in[0,t]} [F_1(s) - F_1(s-)]\,[G_2(s) - G_2(s-)]$

$-F_2(0)\,G_1(0) - \int_0^t F_2dG_1 - \int_0^t G_1dF_2 + \sum_{s\in[0,t]} [F_2(s) - F_2(s-)]\,[G_1(s) - G_1(s-)]$

$= g(0)\,h(0) + \int_0^t gdh + \int_0^t hdg - \sum_{s\in[0,t]} [g(s) - g(s-)]\,[h(s) - h(s-)]$.

■

9.3. Fórmula de cambio de variable

TEOREMA 9.8. *Sea* $g : [0,\infty) \to \mathbb{R}$ *una función continua por la derecha, de variación acotada sobre cualquier intervalo compacto, y* $F : \mathbb{R} \to \mathbb{R}$ *una función de clase* C^1 *, entonces* $F \circ g$ *es de variación acotada sobre cualquier intervalo compacto y:*

$F(g(t)) = F(g(0)) + \int_0^t F'(g(s))dg^c(s) + \sum_{s\in[0,t]} [F(g(s)) - F(g(s-))]$

para cualquier $t \in [0, \infty)$.

Demostración

Para cualquier $t \in [0, \infty)$,se tiene:

$g(t) = g(0) + \int_0^t dg(s) = g(0) + \int_0^t dg^c(s) + \sum_{s \in [0,t]} [g(s) - g(s-)]$.

Supongamos que:

$g^k(t) = g^k(0) + k \int_0^t g^{k-1}(s)\, dg^c(s) + \sum_{s \in [0,t]} \left[g^k(s) - g^k(s-)\right]$

para cualquier $t \in [0, \infty)$, donde $k \in \mathbb{N}$.

Entonces, utilizando la fórmula de integración por partes:

$g^{k+1}(t) = g^{k+1}(0) + \int_0^t g^k dg + \int_0^t g dg^k - \sum_{s \in [0,t]} \left[g^k(s) - g^k(s-)\right] [g(s) - g(s-)]$

$= g^{k+1}(0) + \int_0^t g^k dg^c + \sum_{s \in [0,t]} g^k(s) [g(s) - g(s-)]$

$+ k \int_0^t g^k dg^c + \sum_{s \in [0,t]} g(s) \left[g^k(s) - g^k(s-)\right] - \sum_{s \in [0,t]} \left[g^k(s) - g^k(s-)\right] [g(s) - g(s-)]$

$= g^{k+1}(0) + \int_0^t g^k dg^c + k \int_0^t g^k dg^c + \sum_{s \in [0,t]} g^k(s) [g(s) - g(s-)]$

$+ \sum_{s \in [0,t]} \left[g^k(s) - g^k(s-)\right] g(s-)$

$= g^{k+1}(0) + (k+1) \int_0^t g^k dg^c + \sum_{s \in [0,t]} g^k(s) [g(s) - g(s-)]$

$+ \sum_{s \in [0,t]} [g(s) - g(s-)] \left[\sum_{j=0}^{k-1} g^j(s)\, g^{k-1-j}(s-)\right] g(s-)$

$= g^{k+1}(0) + (k+1) \int_0^t g^k dg^c$

$+ \sum_{s \in [0,t]} [g(s) - g(s-)] \left\{g^k(s) + \left[\sum_{j=0}^{k-1} g^j(s)\, g^{k-1-j}(s-)\right] g(s-)\right\}$

$= g^{k+1}(0) + (k+1) \int_0^t g^k dg^c$

$+ \sum_{s \in [0,t]} [g(s) - g(s-)] \left[g^k(s) + \sum_{j=0}^{k-1} g^j(s)\, g^{k-j}(s-)\right]$

$= g^{k+1}(0) + (k+1) \int_0^t g^k dg^c$

$+ \sum_{s \in [0,t]} [g(s) - g(s-)] \sum_{j=0}^{k} g^j(s)\, g^{k-j}(s-)$

$= g^{k+1}(0) + (k+1) \int_0^t g^k dg^c + \sum_{s \in [0,t]} \left[g^{k+1}(s) - g^{k+1}(s-)\right]$

para cualquier $t \in [0, \infty)$.

Así que, por el principio de inducción matemática:

$g^n(t) = g^n(0) + n\int_0^t g^{n-1}(s)\, dg^c(s) + \sum_{s\in[0,t]} [g^n(s) - g^n(s-)]$

para cualquier $n \in \mathbb{N}$ y cualquier $t \in [0, \infty)$.

Sea $p : \mathbb{R} \to \mathbb{R}$ un polinomio dado por $p(x) = \sum_{k=0}^n a_k x^k$, donde $n \in \mathbb{N}$. Entonces, por la linealidad de la integral:

$p(g(t)) = \sum_{k=0}^n a_k g^k(t)$

$= a_0 + \sum_{k=1}^n a_k \left\{ g^k(0) + k\int_0^t g^{k-1}(s)\, dg^c(s) + \sum_{s\in[0,t]} \left[g^k(s) - g^k(s-)\right] \right\}$

$= \sum_{k=0}^n a_k g^k(t)$

$= a_0 + \sum_{k=1}^n a_k g^k(0) + \sum_{k=1}^n \int_0^t k a_k g^{k-1}(s)\, dg^c(s) + \sum_{k=1}^n \sum_{s\in[0,t]} \left[a_k g^k(s) - a_k g^k(s-)\right]$

$= p(g(0)) + \int_0^t p'(g(s)) dg^c(s) + \sum_{s\in[0,t]} [p(g(s)) - p(g(s-))]$.

Sea $M = \sup\{\max(|g(s)|, |g(s-)|) : s \in [0,t]\}$, tomemos $c > M$ y definamos las funciones $G_1 : \mathbb{R} \to \mathbb{R}$, $G_2 : \mathbb{R} \to \mathbb{R}$ y $G : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:

$G_1(s) = \left[\frac{F(-M)}{(M-c)^2} + \frac{(M-c)F'(-M)+2F(-M)}{(M-c)^3}(s+M)\right](s+c)^2$,

$G_2(s) = \left[\frac{F(M)}{(c-M)^2} + \frac{(c-M)F'(M)+2F(M)}{(c-M)^3}(s-M)\right](s-c)^2$,

$$G(x) = \begin{cases} G_1(x) & \text{si } x \in [-c, -M) \\ F(x) & \text{si } x \in [-M, M] \\ G_2(x) & \text{si } x \in (M, c] \\ 0 & \text{en otro caso} \end{cases}$$

Entonces:

$G_1(-c) = 0$, $G_1(-M) = F(-M)$, $G_1'(-c) = 0$, $G_1'(-M) = F'(-M)$,

$G_2(c) = 0$, $G_2(M) = F(M)$, $G_2'(c) = 0$, $G_2'(M) = F'(M)$,

$G(x) = F(x)$ y $G'(x) = F'(x)$ para cualquier $x \in [-M, M]$.

Además, G es de clase C^1 y nula fuera del intervalo $(-c, c)$, así que existe una sucesión $(p_n)_{n\in\mathbb{N}}$ de polinomios $p_n : \mathbb{R} \to \mathbb{R}$ tales que $(p_n)_{n\in\mathbb{N}}$ y $(p_n')_{n\in\mathbb{N}}$ convergen uniformemente a G y G', respectivamente, en el intervalo $(-c, c)$.

Para cada $n \in \mathbb{N}$, se tiene:

$p_n(g(t)) = p_n(g(0)) + \int_0^t p_n'(g(s)) dg^c(s) + \sum_{s\in[0,t]} [p_n(g(s)) - p_n(g(s-))]$.

En particular, $p_n \circ g$ es de variación acotada sobre cualquier intervalo compacto.

Restringidas al intervalo $[-M, M]$, $p_n' \circ g$ está acotada para cualquier $n \in \mathbb{N}$. Sea entonces $C \in \mathbb{R}$ tal que $|p_n' \circ g| \leq C$ para cualquier $n \in \mathbb{N}$.

Si $F_1 : [0, \infty) \to \mathbb{R}$ y $F_2 : [0, \infty) \to \mathbb{R}$ son dos funciones no decrecientes continuas tales que $g^c = F_1 - F_2$, entonces μ_{F_1} y μ_{F_2} restringidas al intervalo $[0, t]$ son medidas finitas, así que, para cualquier $n \in \mathbb{N}$, $p_n' \circ g$ es integrable con respecto a μ_{F_1} y μ_{F_2}. Por lo tanto, aplicando el teorema de la convergencia dominada:

$\text{lím}_{n \rightsquigarrow \infty} \int_0^t p_n'(g\,(s))dg^c\,(s) = \int_0^t F'(g\,(s))dg^c\,(s)$.

Para cada $n \in \mathbb{N}$, sean $F_1^{(n)} : [0, \infty) \to \mathbb{R}$ y $F_2^{(n)} : [0, \infty) \to \mathbb{R}$ dos funciones no decrecientes continuas por la derecha tales que $p_n \circ g = F_1^{(n)} - F_2^{(n)}$.

Las series $\sum_{s \in [0,t]} \left[F_1^{(n)}\,(s) - F_1^{(n)}\,(s-) \right]$ y $\sum_{s \in [0,t]} \left[F_2^{(n)}\,(s) - F_2^{(n)}\,(s-) \right]$ son absolutamente convergentes ya que sus términos son no negativos y:

$\sum_{s \in [0,t]} \left[F_1^{(n)}\,(s) - F_1^{(n)}\,(s-) \right] \leq F_1^{(n)}\,(t) - F_1^{(n)}\,(0)$,

$\sum_{s \in [0,t]} \left[F_2^{(n)}\,(s) - F_2^{(n)}\,(s-) \right] \leq F_2^{(n)}\,(t) - F_2^{(n)}\,(0)$.

Así que la serie:

$\sum_{s \in [0,t]} [p_n(g\,(s) - p_n(g\,(s-)] = \sum_{s \in [0,t]} \left\{ \left[F_1^{(n)}\,(s) - F_1^{(n)}\,(s-) \right] - \left[F_2^{(n)}\,(s) - F_2^{(n)}\,(s-) \right] \right\}$ es absolutamente convergente ya que:

$\left| \left[F_1^{(n)}\,(s) - F_1^{(n)}\,(s-) \right] - \left[F_2^{(n)}\,(s) - F_2^{(n)}\,(s-) \right] \right|$

$\leq \left[F_1^{(n)}\,(s) - F_1^{(n)}\,(s-) \right] + \left[F_2^{(n)}\,(s) - F_2^{(n)}\,(s-) \right]$.

También la serie $\sum_{s \in [0,t]} |g\,(s) - g\,(s-)|$ es convergente ya que si $F_1^{(0)} : [0, \infty) \to \mathbb{R}$ y $F_2^{(0)} : [0, \infty) \to \mathbb{R}$ son dos funciones no decrecientes continuas por la derecha tales que $g = F_1 - F_2$, entonces:

$\sum_{s \in [0,t]} |g\,(s) - g\,(s-)| = \sum_{s \in [0,t]} |F_1\,(s) - F_1\,(s-) - [F_2\,(s) - F_2\,(s-)]|$

$\leq \sum_{s \in [0,t]} |F_1\,(s) - F_1\,(s-)| + \sum_{s \in [0,t]} |F_2\,(s) - F_2\,(s-)| \leq F_1\,(t) - F_1\,(0) + F_2\,(t) - F_2\,(0) < \infty$.

Además, como, para cualquier $n \in \mathbb{N}$, p_n y F son funciones continuas en el intervalo $[-c, c]$ y derivables con derivada continua en el intervalo $(-c, c)$, para cualquier $s \in [0, t]$ tal que $g\,(s-) \neq g\,(s)$, se tiene:

$(p_n - F)\,(g\,(s)) - (p_n - F)\,(g\,(s-)) = (p_n' - F')\left(\xi_s^{(n)}\right)[g\,(s)) - g\,(s-)]$

donde $\xi_s^{(n)} \in (\min(g(s), g(s-)), \max(g(s), g(s-)))$.

Por lo tanto:

$\left|\sum_{s\in[0,t]} [(p_n - F)(g(s)) - (p_n - F)(g(s-))]\right| = \left|\sum_{s\in[0,t]} (p_n' - F')\left(\xi_s^{(n)}\right)[g(s) - g(s-)]\right|$

$\leq \sum_{s\in[0,t]} \left|(p_n' - F')\left(\xi_s^{(n)}\right)\right| |g(s) - g(s-)|$.

Como la sucesión $(p_n' - F')_{n\in\mathbb{N}}$ converge uniformemente a cero, dada $\varepsilon > 0$ existe $N \in \mathbb{N}$ tal que $|(p_n' - F')(x)| < \frac{1}{\sum_{s\in[0,t]} |g(s)-g(s-)|}\varepsilon$ para cualquier $x \in (-c, c)$ y $n \geq N$, así que:

$\sum_{s\in[0,t]} \left|(p_n' - F')\left(\xi_s^{(n)}\right)\right| |g(s)) - g(s-)| < \varepsilon$

para cualquier $n \geq N$.

Por lo tanto:

$\lim_{n\rightsquigarrow\infty} \sum_{s\in[0,t]} [(p_n - F)(g(s)) - (p_n - F)(g(s-))] = 0$.

Así que, tomando límites cuando $n \rightsquigarrow \infty$ en la expresión:

$p_n(g(t)) = p_n(g(0)) + \int_0^t p_n'(g(s))dg^c(s) + \sum_{s\in[0,t]} [p_n(g(s)) - p_n(g(s-))]$,

se obtiene:

$F(g(t)) = F(g(0)) + \int_0^t F'(g(s))dg^c(s) + \sum_{s\in[0,t]} [F(g(s)) - F(g(s-))]$

para cualquier $t \in [0, \infty)$.

■

CAPÍTULO 10

CONVERGENCIA

10.1. Introducción

En este capítulo analizaremos diferentes tipos de convergencia de una sucesión de funciones medibles definidas sobre un espacio de medida $(\mathbb{F}, \Im, \mu)$ y con valores en $\overline{\mathbb{R}}$.

En todo el capítulo, $(\mathbb{F}, \Im, \mu)$ será un espacio de medida completo. La medibilidad de una función $f : \mathbb{F} \to \mathbb{R}$ (resp. $f : \mathbb{F} \to \overline{\mathbb{R}}$) será entendida considerando sobre $\mathbb{R}$ (resp. $\overline{\mathbb{R}}$) la σ-álgebra de los conjuntos borelianos en $\mathbb{R}$ (resp. $\overline{\mathbb{R}}$).

Desde los inicios del Cálculo Diferencial e Integral se planteó el problema de expresar una función como una serie de funciones simples (recordemos, por ejemplo, el desarrollo de una función en serie de Taylor). Más adelante, se planteó el problema de expresar una función como una serie trigonométrica. En particular, recordemos que Fourier, en el año 1822, afirmó que una función arbitraria f, definida y acotada en el intervalo $[-L, L]$, puede representarse mediante la siguiente serie trigonométrica:

$$f(x) = \tfrac{1}{2}a_0 + \textstyle\sum_{n=1}^{\infty} \left[a_n \cos \tfrac{n\pi x}{L} + b_n sen \tfrac{n\pi x}{L}\right],$$

donde $a_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx$ y, para cada $n \in \mathbb{N}$:

$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{n\pi x}{L}dx$,

$b_n = \frac{1}{L}\int_{-L}^{L} f(x) sen\frac{n\pi x}{L}dx$.

Esta afirmación de Fourier condujo a investigar a fondo cuándo la serie:

$$\tfrac{1}{2}a_0 + \textstyle\sum_{n=1}^{\infty} \left[a_n \cos \tfrac{n\pi x}{L} + b_n sen \tfrac{n\pi x}{L}\right],$$

con los valores de los coeficientes dados por Fourier, converge efectivamente a la función f.

Las investigaciones alrededor del problema de la convergencia de una serie de funciones condujeron a la necesidad de profundizar, en general, en el problema de la convergencia de una sucesión de funciones. Surgió en este proceso la definición de la convergencia uniforme; esto a raíz de que Cauchy afirmaba que si los términos de una serie son funciones continuas y la serie converge, entonces la función a la cual converge es continua. N. H. Abel hizo ver que esta afirmación no es válida en general y Weierstrass, introduciendo el concepto de convergencia uniforme, demostró que la afirmación de Cauchy es válida si la convergencia de la serie es uniforme.

Otro problema relacionado con las series de funciones era el determinar bajo que condiciones se puede integrar término a término una serie convergente de funciones para obtener la integral de la función a la cual converge.

Al desarrollar Lebesgue su teoría de integración, uno de los aspectos centrales de su planteamiento fue el obtener una definición de integral con la propiedad de que si se tiene una sucesión convergente de funciones integrables, la función límite sea integrable y su integral sea igual al límite de la sucesión formada por las integrales de las funciones que componen la sucesión dada. Como vimos en el capítulo anterior, la integral que definió satisface esta propiedad bajo condiciones bastante generales.

La integral que definió Lebesgue tiene la característica de que al modificar los valores de una función integrable en los puntos de un conjunto de medida cero, la función sigue siendo integrable y la integral de la función modificada es igual a la integral de la función original. Es decir, para fines de la integración de funciones, los conjuntos de medida cero son despreciables. En particular, en lo que respecta a la convergencia de una sucesión de funciones, esto lleva a que no es necesario tratar con sucesiones de funciones que converjan en todos los puntos, puede uno limitarse a la convergencia fuera de un conjunto de medida cero. Surgió así el concepto de convergencia casi en todas partes.

De esta forma, en el desarrollo de la teoría de integración, incluyendo los trabajos anteriores al de Lebesgue, se fueron encontrando diferentes tipos de convergencia de una sucesión de funciones.

Un tipo de convergencia que surgió en este contexto es la convergencia en medida. Dentro de la teoría de integración la idea surgió del problema de la integración término a término de una serie de funciones, buscando condiciones menos restrictivas que la convergencia uniforme de la serie para asegurar que se puede integrar término a término una serie convergente de funciones para obtener la integral de la función a la cual converge. Sin definirla explícitamente, este tipo de convergencia se encuentra formulada en un artículo de L. Kronecker del año 1878, obviamente sin referirse al concepto de medida, que aún no se había formulado, sino al de contenido. La definición explícita fue dada en el año 1909 por F. Riesz en un artículo titulado *Sur les suites des fonctions mesurables* ([**81**]), donde, entre otras cosas, demostró que si una sucesión de funciones medibles converge en medida a una función medible, entonces existe una subsucesión que converge casi en todas partes.

Sin embargo, la convergencia en medida tiene una historia más antigua; sin definirse explícitamente, era utilizada en el CálculoTeoría de Probabilidades desde la publicación del teorema de Bernoulli en el año 1713 ([**4**]), el cual establece la convergencia, que hoy se denomina en medida, de una determinada sucesión de variables aleatorias. Este resultado de Bernoulli marcó la pauta para el desarrollo del CálculoTeoría de Probabilidades hasta principios del siglo XX, cuando se llegó a la formulación general de los llamados teoremas límite, entre los cuales se encuentra la ley débil de los grandes números, la cual es una generalización del resultado de Bernoulli.

También en el contexto de la Teoría de la Probabilidad surgió otro tipo de convergencia llamada en distribución, la cual proviene del teorema que demostró de Moivre en el año 1733 ([**28**]) y que en su forma general se le conoce como Teorema Central del Límite.

En 1906, Hilbert introdujo el espacio l^2, el cual está formado por las sucesiones $(x_n)_{n\in\mathbb{N}}$ tales que la serie $\sum_{n=1}^{\infty} x_n^2$ es convergente, y definió la distancia entre dos elementos $x = (x_n)_{n\in\mathbb{N}}$ y $y = (y_n)_{n\in\mathbb{N}}$ de l^2 de la siguiente manera:

$$d(x, y) = \left(\sum_{n=1}^{\infty} (x_n - y_n)^2\right)^{\frac{1}{2}}.$$

En 1907 ([**78**]), Riesz demostró que existe un isomorfismo entre el espacio l^2 y el espacio de funciones medibles $f : [a, b] \to \mathbb{R}$ tales que f^2 es integrable. Previamente, en el año 1906, Hilbert había introducido el concepto de funciones ortogonales: dos funciones continuas f y g, definidas sobre el intervalo $[a, b]$, son ortogonales si $\int_a^b f(x)\, g(x)\, dx = 0$; un sistema ortogonal es una familia de funciones ortogonales por parejas y tales que la integral del cuadrado de cada una de ellas es igual a 1 (actualmente, cuando se agrega la segunda condición, se le llama sistema ortonormal).

Riesz estableció el isomorfismo mencionado con el siguiente resultado:

Sea $(\varphi_k)_{k\in\mathbb{N}}$ un conjunto ortogonal de funciones y $(c_k)_{k\in\mathbb{N}}$ una sucesión de números reales, entonces, existe una función medible f , de cuadrado integrable, tal que

$c_k = \int_a^b f(x)\, \varphi_k(x)\, dx$ para cualquier $k \in \mathbb{N}$ si y sólo si la serie $\sum_{k=1}^{\infty} c_k^2$ converge.

En el mismo año, Ernst Fischer ([**34**]) demostró que el conjunto de funciones medibles de cuadrado integrable es completo con la métrica dada por $d(f, g) = \left(\int_a^b (f(x) - g(x))^2\, dx\right)^{\frac{1}{2}}$. Más tarde Riesz daría su propia demostración de este resultado. El concepto de espacio métrico había sido introducido por Fréchet en 1906 en su tesis doctoral ([**36**]).

También en 1907 ([**79**]), Riesz demostró que si U es una función lineal y continua, definida sobre el conjunto de funciones medibles de cuadrado integrable, entonces existe una función medible g, de cuadrado integrable, tal que $U(f) = \int_a^b f(x)\, g(x)\, dx$ para cualquier función f medible de cuadrado integrable. El mismo resultado fue demostrado en el mismo año, de manera independiente, por Maurice Fréchet.

Unos años después, en 1909 ([**81**]), Riesz definió el espacio $L^p([a,b])$, para $p \in (1,\infty)$, como el conjunto de funciones $f : [a,b] \to \mathbb{R}$ tales que $\int_a^b |f(x)|^p \, dx < \infty$ (identificando aquellas que son iguales excepto en un conjunto de medida cero) y demostró que dicho espacio es completo con la métrica dada por $d(f,g) = \left(\int_a^b |f(x) - g(x)|^p \, dx \right)^{\frac{1}{p}}$. Surgió así otro tipo de convergencia de una sucesión de funciones medibles: la convergencia en L^p.

10.2. Convergencia casi en todas partes

Definición 10.1. *Diremos que una sucesión de funciones medibles $\{f_n\}_{n\in\mathbb{N}}$ converge casi en todas partes a una función medible f si $\lim_{n\rightsquigarrow\infty} f_n = f$ excepto a lo más en un conjunto de medida cero. Si éste es el caso, se escribirá $f_n \xrightarrow{c.t.p.} f$.*

Las siguientes propiedades se siguen inmediatamente de las correspondientes propiedades para las sucesiones de números reales:

(i) Si una sucesión $(f_n)_{n\in\mathbb{N}}$ converge casi en todas partes a f, entonces cualquier subsucesión de $(f_n)_{n\in\mathbb{N}}$ también converge casi en todas partes a f.
(ii) Si $\{f_n\}$ es una sucesión de funciones medibles tal que $f_n \xrightarrow{c.t.p.} f$ y $f_n \xrightarrow{c.t.p.} g$, entonces $f = g$ casi en todas partes.
(iii) Si c es una constante y $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{c.t.p.} f$, entonces $cf_n \xrightarrow{c.t.p.} cf$.
(iv) Si $(f_n)_{n\in\mathbb{N}}$ y $(g_n)_{n\in\mathbb{N}}$ son dos sucesiones de funciones medibles tales que $f_n \xrightarrow{c.t.p.} f$ y $g_n \xrightarrow{c.t.p.} g$, entonces $f_n + g_n \xrightarrow{c.t.p.} f + g$ y $f_n g_n \xrightarrow{c.t.p.} fg$.

Los resultados que siguen, con relación a la convergencia casi en todas partes, se entienden mejor si se tienen en mente los siguientes conceptos:

Dada una sucesión $A_1, A_2, \ldots$ de subconjuntos de un conjunto Ω, se define el límite inferior (lím ínf) y el límite superior (lím sup) de esa sucesión de la siguiente manera:

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m,$$

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

Obsérvese que $\liminf A_n$ está formado por todos los elementos $x \in \mathbb{F}$ para los cuales existe $n \in \mathbb{N}$ tal que $x \in A_m$ para cualquier $m \geq n$, mientras que $\limsup A_n$ está formado por todos los elementos $x \in \mathbb{F}$ que pertenecen a una infinidad de conjuntos de la sucesión. Así que se tiene siempre $\liminf A_n \subset \limsup A_n$.

Si $\liminf A_n = \limsup A_n$, se dice que la sucesión $(A_n)_{n\in\mathbb{N}}$ converge y al valor común de $\liminf A_n$ y $\limsup A_n$ se le llama $\lim A_n$.

Por ejemplo, si $(A_n)_{n\in\mathbb{N}}$ es una sucesión creciente (resp. decreciente) de subconjuntos de Ω, entonces converge y $\lim A_n = \cup_{n=1}^{\infty} A_n$ (resp. $\lim A_n = \cap_{n=1}^{\infty} A_n$).

Teorema 10.1. *Supongamos que μ es finita y sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles. Entonces $f_n \xrightarrow{c.t.p.} 0$ si y sólo si para cualquier $\varepsilon > 0$, se tiene:*

$\mu\left(\limsup\left\{y \in \mathbb{F} : |f_n(y)| > \varepsilon\right\}\right) = 0.$

Demostración

Supongamos primero que $\lim_{n\rightsquigarrow\infty} f_n = 0$ casi en todas partes. Entonces existe un conjunto $E_0 \subset \mathbb{F}$ de medida 0 tal que si $x \in E_0^c$ entonces $\lim_{n\rightsquigarrow\infty} f_n(x) = 0$. Así que, dado $x \in E_0^c$ y $\varepsilon > 0$, existe N tal que $|f_n(x)| \leq \varepsilon$ para cualquier $n \geq N$, esto significa que $x \in \bigcap_{n=N}^{\infty}\left\{y \in \mathbb{F} : |f_n(y)| \leq \varepsilon\right\}$.

Dicho de otra forma, si $x \in E_0^c$, entonces, dada cualquier $\varepsilon > 0$:

$x \in \bigcup_{m=1}^{\infty}\left[\bigcap_{n=m}^{\infty}\left\{y \in \mathbb{F} : |f_n(y)| \leq \varepsilon\right\}\right].$

Así que:

$E_0^c \subset \bigcup_{m=1}^{\infty}\left[\bigcap_{n=m}^{\infty}\left\{x \in \mathbb{F} : |f_n(x)| \leq \varepsilon\right\}\right].$

Por lo tanto:

$\mu\left(\bigcap_{m=1}^{\infty}\left[\bigcup_{n=m}^{\infty}\left\{y \in \mathbb{F} : |f_n(y)| > \varepsilon\right\}\right]\right) = 0.$

Inversamente, supongamos que, para cualquier $\varepsilon > 0$, se tiene:

$\mu\left(\bigcap_{m=1}^{\infty}\left[\bigcup_{n=m}^{\infty}\left\{y \in \mathbb{F} : |f_n(y)| > \varepsilon\right\}\right]\right) = 0.$

Para cada $r \in \mathbb{N}$, sea:

$B_r = \bigcap_{m=1}^{\infty}\left[\bigcup_{n=m}^{\infty}\left\{y \in \mathbb{F} : |f_n(y)| > \frac{1}{r}\right\}\right].$

Se tiene $\mu(B_r) = 0$ para cualquier $r \in \mathbb{N}$ y la sucesión de eventos $B_1, B_2, \ldots$ es creciente, así que:

$\mu\left(\bigcup_{r=1}^{\infty} B_r\right) = \lim_{r\rightsquigarrow\infty} \mu(B_r) = 0.$

Pero, $B_r^c = \left\{x \in \mathbb{F} : \text{Existe } N(x) \text{ tal que } |f_n(x)| \leq \frac{1}{r} \text{ para cualquier } n \geq N(x)\right\}$. De manera que si $x \in \bigcap_{r=1}^{\infty} B_r^c$, entonces para cualquier $r \in \mathbb{N}$ existe $N(x)$ tal que $|f_n(x)| \leq \frac{1}{r}$ para cualquier $n \geq N(x)$. En particular, dada $\varepsilon > 0$ sea $r \in \mathbb{N}$ tal que $\frac{1}{r} < \varepsilon$ y $N(x)$ tal que $|f_n(x)| \leq \frac{1}{r}$ para cualquier $n \geq N(x)$, entonces $|f_n(x)| < \varepsilon$ para cualquier $n \geq N(x)$, lo cual significa que $\lim_{n\rightsquigarrow\infty} f_n(x) = 0$. Es decir:

$\bigcap_{r=1}^{\infty} B_r^c \subset [\lim_{n\rightsquigarrow\infty} f_n = 0].$

Sea $E_0 = \bigcup_{r=1}^{\infty} B_r$. Entonces $\mu(E_0) = 0$ y si $x \in E_0^c = \bigcap_{r=1}^{\infty} B_r^c$, entonces $\lim_{n \rightsquigarrow \infty} f_n(x) = 0$. Así que $\lim_{n \to \infty} f_n = 0$ casi en todas partes.

■

Corolario 10.1. *Supongamos que μ es finita y sea $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles. Entonces $f_n \xrightarrow{c.t.p.} f$ si y sólo si para cualquier $\varepsilon > 0$, se tiene:*

$\mu(\limsup \{y \in \mathbb{F} : |f_n(y) - f(y)| > \varepsilon\}) = 0.$

Teorema 10.2 (Lema de Borel-Cantelli). *Sea $E_1, E_2, \ldots$ una sucesión de conjuntos medibles tales que $\sum_{n=1}^{\infty} \mu(E_n) < \infty$, entonces:*

$\mu(\limsup E_n) = 0.$

Demostración

Sea $B = \limsup E_n$.

Para cada $m \in \mathbb{N}$, sea $B_m = \bigcup_{n=m}^{\infty} E_n$. Entonces la sucesión de eventos B_m es decreciente y $B = \bigcap_{m=1}^{\infty} B_m$, así que:

$\mu(B) = \mu\left[\bigcap_{m=1}^{\infty} B_m\right] = \lim_{m \rightsquigarrow \infty} \mu\left[\bigcup_{n=m}^{\infty} E_n\right] \leq \lim_{m \rightsquigarrow \infty} \sum_{n=m}^{\infty} \mu(E_n) = 0.$

■

Corolario 10.2. *Sea $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tales que:*

$\sum_{n=1}^{\infty} \mu[|f_n| > \varepsilon] < \infty$ *para cualquier* $\varepsilon > 0$.

Entonces $f_n \xrightarrow{c.t.p.} 0$.

Demostración

Dada $\varepsilon > 0$, sea $A(\varepsilon) = \limsup \{y \in \mathbb{F} : |f_n(y)| > \varepsilon\}$.

Por la proposición 10.2, $\mu[A(\varepsilon)] = 0$ para cualquier $\varepsilon > 0$. Así que el resultado se sigue aplicando el corolario 10.1.

■

Corolario 10.3. *Sea f una función medible y $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tales que $\sum_{n=1}^{\infty} \mu[|f_n - f| > \varepsilon] < \infty$ para cualquier $\varepsilon > 0$. Entonces $f_n \xrightarrow{c.t.p.} f$.*

10.3. Convergencia en medida

Definición 10.2. *Diremos que una sucesión de funciones medibles $(f_n)_{n \in \mathbb{N}}$ converge en medida si existe una función medible f tal que:*

$\lim_{n \rightsquigarrow \infty} \mu(\{y \in \mathbb{F} : |f_n(y) - f(y)| > \varepsilon\}) = 0$

para cualquier $\varepsilon > 0$. *Si éste es el caso, se escribirá* $f_n \xrightarrow{\mu} f$.

Obviamente si una sucesión $(f_n)_{n\in\mathbb{N}}$ converge en medida a f, entonces cualquier subsucesión de $(f_n)_{n\in\mathbb{N}}$ también converge en medida a f.

PROPOSICIÓN 10.1. *Sea* $\{f_n\}$ *una sucesión de funciones medibles tal que* $f_n \xrightarrow{\mu} f$ *y* $f_n \xrightarrow{\mu} g$, *entonces* $f = g$ *casi en todas partes.*

Demostración

Como $|f-g| \leq |f_n - f| + |f_n - g|$, entonces:

$[|f-g| > \varepsilon] \subset [|f_n - f| + |f_n - g| > \varepsilon]$.

Además, para cualquier $\varepsilon > 0$, se tiene:

$[|f_n - f| + |f_n - g| > \varepsilon] \subset \left[|f_n - f| > \frac{\varepsilon}{2}\right] \cup \left[|f_n - g| > \frac{\varepsilon}{2}\right]$.

Por lo tanto:

$\mu\,[|f-g| > \varepsilon] \leq \mu\left[|f_n - f| > \frac{\varepsilon}{2}\right] + \mu\left[|f_n - g| > \frac{\varepsilon}{2}\right]$.

Así que, tomando límites, se obtiene $\mu\,[|f-g| > \varepsilon] = 0$ para cualquier $\varepsilon > 0$.

Finalmente, $[|f-g| > 0] = \bigcup_{n=1}^{\infty}\left[|f-g| > \frac{1}{n}\right]$, así que:

$\mu\,[|f-g| > 0] \leq \sum_{n=1}^{\infty}\mu\left[|f-g| > \frac{1}{n}\right] = 0$.

■

PROPOSICIÓN 10.2. *Sea* c *una constante y* $(f_n)_{n\in\mathbb{N}}$ *una sucesión de funciones medibles tal que* $f_n \xrightarrow{\mu} f$, *entonces* $cf_n \xrightarrow{\mu} cf$.

Demostración

$\text{lím}_{n\rightsquigarrow\infty}\,\mu\,[|cf_n - cf| > \varepsilon] = \text{lím}_{n\rightsquigarrow\infty}\,\mu\left[|f_n - f| > \frac{\varepsilon}{|c|}\right] = 0$.

■

PROPOSICIÓN 10.3. *Sean* $(f_n)_{n\in\mathbb{N}}$ *y* $(g_n)_{n\in\mathbb{N}}$ *dos sucesiones de funciones medibles tales que* $f_n \xrightarrow{\mu} f$ *y* $g_n \xrightarrow{\mu} g$, *entonces* $f_n + g_n \xrightarrow{\mu} f + g$.

Demostración

Como $|f_n - f + g_n - g| \leq |f_n - f| + |g_n - g|$, se tiene:

$[|f_n - f + g_n - g| > \varepsilon] \subset \left[|f_n - f| > \frac{\varepsilon}{2}\right] \cup \left[|g_n - g| > \frac{\varepsilon}{2}\right]$.

Así que:

$\text{lím}_{n\rightsquigarrow\infty}\,\mu\,[|f_n + g_n - f - g| > \varepsilon]$

$\leq \text{lím}_{n \rightsquigarrow \infty} \mu \left[|f_n - f| > \frac{\varepsilon}{2} \right] + \text{lím}_{n \rightsquigarrow \infty} \mu \left[|g_n - g| > \frac{\varepsilon}{2} \right] = 0.$

■

PROPOSICIÓN 10.4. *Sean f una función medible, $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tales que $f_n \xrightarrow{\mu} f$ y $g : \mathbb{R} \to \mathbb{R}$ una función uniformemente continua, entonces $g \circ f_n \xrightarrow{\mu} g \circ f$.*

Demostración

Como g es uniformemente continua en $\mathbb{R}$, dada $\varepsilon > 0$ existe $\delta > 0$ tal que $|g(u) - g(v)| < \varepsilon$ para cualesquiera $u, v \in \mathbb{R}$ tales que $|v - u| < \delta$. Así que:

$|g \circ f_n(x) - g \circ f(x)| < \varepsilon$

para cualquier $x \in \mathbb{F}$ tal que $|f_n(x) - f(x)| < \delta$.

Por lo tanto:

$\text{lím sup}_{n \rightsquigarrow \infty} \mu (\{y \in \mathbb{F} : |g \circ f_n(y) - g \circ f(y)| > \varepsilon\})$

$\leq \text{lím sup}_{n \rightsquigarrow \infty} \mu (\{y \in \mathbb{F} : |g \circ f_n(y) - g \circ f(y)| \geq \varepsilon\})$

$\leq \text{lím sup}_{n \rightsquigarrow \infty} \mu (\{y \in \mathbb{F} : |f_n(y) - f(y)| \geq \delta\})$

$\leq \text{lím}_{n \rightsquigarrow \infty} \mu \left(\left\{ y \in \mathbb{F} : |f_n(y) - f(y)| > \frac{1}{2}\delta \right\} \right) = 0.$

Así que $g \circ f_n \xrightarrow{\mu} g \circ f$.

■

COROLARIO 10.4. *Sean f una función medible, $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tales que $f_n \xrightarrow{\mu} f$ y $g : \mathbb{R} \to \mathbb{R}$ una función continua nula fuera de un intervalo $[a, b]$, entonces $g \circ f_n \xrightarrow{\mu} g \circ f$.*

Demostración

Como g es uniformemente continua en $[a, b]$ y nula fuera de ese intervalo, es uniformemente continua en $\mathbb{R}$.

■

TEOREMA 10.3. *Supongamos que μ es finita y sean f una función medible, $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{\mu} f$ y $g : \mathbb{R} \to \mathbb{R}$ una función continua, entonces $g \circ f_n \xrightarrow{\mu} g \circ f$.*

Demostración

Como $\sum_{k=0}^{\infty} \mu [k \leq |f| < k + 1] = \mu(\mathbb{F}) < \infty$, entonces dada $\eta > 0$ existe M tal que:

$\mu [|f| > M] \leq \mu [|f| \geq M] = \sum_{k=M}^{\infty} P [k \leq |f| < k + 1] < \frac{1}{2}\eta.$

También, como $f_n \xrightarrow{\mu} f$, existe N tal que, si $n \geq N$, entonces:

$\mu\left[|f_n - f| > M\right] < \frac{1}{2}\eta$.

Sea $D_n = \{y \in \mathbb{F} : |f(y)| \leq M\} \cap \{y \in \mathbb{F} : |f_n(y) - f(y)| \leq M\}$.

Entonces:

$D_n^c = \{y \in \mathbb{F} : |f(y)| > M\} \cup \{y \in \mathbb{F} : |f_n(y) - f(y)| > M\}$.

Así que:

$\mu(D_n^c) < \eta$.

Si $y \in D_n$, entonces:

$|f_n(y)| \leq |f_n(y) - f(y)| + |f(y)| \leq 2M$.

Definamos:

$$g^{(M)}(u) = \begin{cases} g(u) & \text{si } |u| < 2M \\ g(2M) & \text{si } u \geq 2M \\ g(-2M) & \text{si } u \leq -2M \end{cases}$$

$g^{(M)}$ es uniformemente continua en $\mathbb{R}$, así que, dada $\varepsilon > 0$ existe $\delta > 0$ tal que:

$\left|g^{(M)}(u) - g^{(M)}(v)\right| < \varepsilon$

para cualesquiera $u, v \in \mathbb{R}$ tales que $|v - u| < \delta$.

Así que:

$\left|g^{(M)} \circ f_n(x) - g^{(M)} \circ f(x)\right| < \varepsilon$

para cualquier $x \in \mathbb{F}$ tal que $|f_n(x) - f(x)| < \delta$.

Por lo tanto, si $n \geq N$, se tiene:

$\mu(\{y \in \mathbb{F} : |g \circ f_n(y) - g \circ f(y)| > \varepsilon\})$

$= \mu(D_n \cap \{y \in \mathbb{F} : |g \circ f_n(y) - g \circ f(y)| > \varepsilon\}) + \mu(D_n^c \cap \{y \in \mathbb{F} : |g \circ f_n(y) - g \circ f(y)| > \varepsilon\})$

$\leq \mu\left(D_n \cap \left\{y \in \mathbb{F} : \left|g^{(M)} \circ f_n(y) - g^{(M)} \circ f(y)\right| \geq \varepsilon\right\}\right) + \mu(D_n^c)$

$\leq \mu\left(\left\{y \in \mathbb{F} : \left|g^{(M)} \circ f_n(y) - g^{(M)} \circ f(y)\right| \geq \varepsilon\right\}\right) + \mu(D_n^c)$

$\leq \mu(\{y \in \mathbb{F} : |f_n(y) - f(y)| \geq \delta\}) + \mu(D_n^c)$

$\leq \mu\left(\left\{y \in \mathbb{F} : |f_n(y) - f(y)| > \frac{1}{2}\delta\right\}\right) + \eta$.

Así que:

$\lim\sup_{n\rightsquigarrow\infty} \mu\left(\{y \in \mathbb{F} : |g \circ f_n(y) - g \circ f(y)| > \varepsilon\}\right)$

$\leq \lim_{n\rightsquigarrow\infty} \mu\left(\left\{y \in \mathbb{F} : |f_n(y) - f(y)| > \frac{1}{2}\delta\right\}\right) + \eta = \eta.$

Como lo anterior es válido para cualquier $\eta > 0$, se concluye que:

$\lim_{n\rightsquigarrow\infty} \mu\left(\{y \in \mathbb{F} : |g \circ f_n(y) - g \circ f(y)| > \varepsilon\}\right) = 0.$

Por lo tanto, $g \circ f_n \xrightarrow{\mu} g \circ f$.

■

Corolario 10.5. *Supongamos que μ es finita y sean f y g dos funciones medibles, y $(f_n)_{n\in\mathbb{N}}$ y $(g_n)_{n\in\mathbb{N}}$ dos sucesiones de funciones medibles tales que $f_n \xrightarrow{\mu} f$ y $g_n \xrightarrow{\mu} g$, entonces $f_n g_n \xrightarrow{\mu} fg$.*

Demostración

Como $f_n g_n = \frac{1}{4}\left[(f_n + g_n)^2 - (f_n - g_n)^2\right]$, entonces:

$f_n g_n \xrightarrow{\mu} \frac{1}{4}\left[(f + g)^2 - (f - g)^2\right] = fg.$

■

Teorema 10.4. *Supongamos que μ es finita y sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{c.t.p.} 0$, entonces $f_n \xrightarrow{\mu} 0$.*

Demostración

Sea $B_m(\varepsilon) = \bigcup_{n=m}^{\infty} \{y \in \mathbb{F} : |f_n(y)| > \varepsilon\}$, entonces la sucesión de eventos $B_1(\varepsilon), B_2(\varepsilon), \ldots$ es decreciente, así que:

$\lim_{m\rightsquigarrow\infty} \mu\left[B_m(\varepsilon)\right] = \mu\left(\bigcap_{m=1}^{\infty} B_m(\varepsilon)\right) = 0.$

Pero $[|f_m| > \varepsilon] \subset B_m(\varepsilon)$. Por lo tanto:

$\lim_{m\rightsquigarrow\infty} \mu\left[|f_m| > \varepsilon\right] \leq \lim_{m\rightsquigarrow\infty} \mu\left[B_m(\varepsilon)\right] = 0.$

■

Corolario 10.6. *Supongamos que μ es finita y sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{c.t.p.} f$, entonces $f_n \xrightarrow{\mu} f$.*

Si μ no es finita, el corolario anterior no es válido en general. En efecto consideremos el ejemplo siguiente:

Ejemplo 10.1. *Sea $\mathbb{F} = \mathbb{R}$, μ la medida de Lebesgue sobre $\mathbb{R}$ y $(f_n)_{n\in\mathbb{N}}$ la sucesión de definidas por $f_n = I_{(n,n+1)}$ para cualquier $n \in \mathbb{N}$. Entonces la sucesión $(f_n(x))_{n\in\mathbb{N}}$ converge a 0 para cualquier $x \in \mathbb{R}$, pero dada $\varepsilon \in (0, 1)$, se tiene:*

$\mu(\{y \in \mathbb{R} : |f_n(y)| > \varepsilon\}) = 1$

para cualquier $n \in \mathbb{N}$.

Así que $(f_n)_{n\in\mathbb{N}}$ *no converge en medida a la función identicamente cero. Más aún,* $(f_n)_{n\in\mathbb{N}}$ *no converge en medida. En efecto, supongamos que* $(f_n)_{n\in\mathbb{N}}$ *converge en medida a la función medible* f*, entonces existe una subsucesión que converge casi en todas partes, así que* f *es la función identicamente cero.*

Como se muestra en el siguiente ejemplo, el inverso del corolario 10.6 no es válido en general.

EJEMPLO 10.2. *Sea* $\mathbb{F} = (0, 1]$*,* μ *la medida de Lebesgue sobre* $\mathbb{F}$ *e* $(I_n)_{n\in\mathbb{N}}$ *la sucesión de intervalos* $J_1 = (0, 1]$*,* $J_2 = (0, \frac{1}{2}]$*,* $J_3 = (\frac{1}{2}, \frac{1}{2}]$*,* $J_4 = (0, \frac{1}{2^2}]$*,* $J_5 = (\frac{1}{2^2}, \frac{2}{2^2}]$*,* $J_6 = (\frac{2}{2^2}, \frac{3}{2^2}]$*,* $J_7 = (\frac{3}{2^2}, \frac{4}{2^2}]$*, ...; es decir, para* $n \in \{0, 1, 2, \ldots\}$ *y* $j \in \{0, 1, 2, \ldots, 2^n - 1\}$*,* $J_{2^n+j} = (\frac{j}{2^n}, \frac{j+1}{2^n}]$*. Para cada* $n \in \mathbb{N}$*, definamos* $f_n = I_{J_n}$*; es decir, para* $n \in \{0, 1, 2, \ldots\}$ *y* $j \in \{0, 1, 2, \ldots, 2^n - 1\}$*,* $f_{2^n+j} = I_{(\frac{j}{2^n}, \frac{j+1}{2^n}]}$*.*

Si $x \in \mathbb{F}$ *y* $n \in \{0, 1, 2, \ldots\}$*, entonces existe un único elemento* $j_0 \in \{0, 1, 2, \ldots, 2^n - 1\}$ *tal que* $x \in (\frac{j}{2^n}, \frac{j+1}{2^n}]$*, así que* $f_{2^n+j_0}(x) = 1$ *y* $f_{2^n+j}(x) = 0$ *para cualquier* $j \in \{0, 1, 2, \ldots, 2^n - 1\} - \{j_0\}$*. Por lo tanto,* $f_n(x) = 1$ *para una infinidad de valores de* n *y* $f_n(x) = 0$ *para una infinidad de valores de* n*. Así que la sucesión* $(f_n(x))_{n\in\mathbb{N}}$ *no converge para ninguna* $x \in \mathbb{F}$*.*

Sin embargo, para cualquier $n \in \{0, 1, 2, \ldots\}$ *y* $j \in \{0, 1, 2, \ldots, 2^n - 1\}$*,* f_{2^n+j} *toma únicamente los valores* 0 *y* 1*, y* $\mu(\{y \in \mathbb{F} : f_{2^n+j}(y) = 1\}) = \frac{1}{2^n}$*. Así que* $\lim_{n\rightsquigarrow\infty} \mu[|f_n| > \varepsilon] = 0$ *para cualquier* $\varepsilon > 0$*. Por lo tanto,* $f_n \xrightarrow{\mu} 0$*.*

De la definición de convergencia en medida podemos ver que, en un sentido, las funciones f_n, para n suficientemente grande, están cercanas a la función límite f. En efecto, si $f_n \xrightarrow{\mu} f$, dada cualquier $\varepsilon > 0$, por pequeña que sea, y dada $\delta > 0$, por pequeña que sea, existe $N \in \mathbb{N}$ tal que:

$$\mu(\{y \in \mathbb{F} : |f_n(y) - f(y)| > \varepsilon\}) < \delta$$

para cualquier $n \geq N$.

Es decir, denotando por A_n al conjunto $\{y \in \mathbb{F} : |f_n(y) - f(y)| > \varepsilon\}$, entonces, para cualquier $n \geq N$, $|f_n(y) - f(y)| \leq \varepsilon$ para cualquier $y \in A_n^c$ y $\mu(A_n) < \delta$.

Lo anterior podría dar la idea de que, fuera de un conjunto de medida pequeña, se tiene convergencia uniforme de la sucesión $(f_n)_{n\in\mathbb{N}}$. Sin embargo, esto no es así ya que el conjunto de medida pequeña (A_n) no es fijo, depende de n. Esto se ve claro en el ejemplo 10.2 ya que, dada $\varepsilon \in (0, 1)$ y $\delta > 0$, digamos $\delta = \frac{1}{2^{n_0}}$, con n_0 grande para asegurar que δ es pequeña, se tiene que, para $n \geq 2^{n_0+1}$, $\mu(A_n) < \delta$, pero:

$A_{2^{n_0+1}} = I_{(\frac{0}{2^{n_0+1}}, \frac{1}{2^{n_0+1}}]}$,

$A_{2^{n_0+1}+1} = I_{\left(\frac{1}{2^{n_0+1}}, \frac{2}{2^{n_0+1}}\right]}$,

...

$A_{2^{n_0+2}-1} = I_{\left(\frac{2^{n_0+1}-1}{2^{n_0+1}}, \frac{2^{n_0+1}}{2^{n_0+1}}\right]}$,

$A_{2^{n_0+2}} = I_{\left(\frac{0}{2^{n_0+2}}, \frac{1}{2^{n_0+2}}\right]}$,

$A_{2^{n_0+2}+1} = I_{\left(\frac{1}{2^{n_0+2}}, \frac{2}{2^{n_0+2}}\right]}$,

...

$A_{2^{n_0+3}-1} = I_{\left(\frac{2^{n_0+2}-1}{2^{n_0+2}}, \frac{2^{n_0+2}}{2^{n_0+1}}\right]}$,

...

Se tiene la convergencia uniforme fuera de un conjunto fijo de medida pequeña, pero únicamente para alguna subsucesión de $(f_n)_{n\in\mathbb{N}}$, lo cual demostramos más adelante.

DEFINICIÓN 10.3. *Diremos que una sucesión de funciones medibles $(f_n)_{n\in\mathbb{N}}$ es de Cauchy en medida si, para cualquier $\varepsilon > 0$ y cualquier $\delta > 0$ existe $N \in \mathbb{N}$ tal que:*

$\mu\left(\{y \in \mathbb{F} : |f_n(y) - f_m(y)| > \varepsilon\}\right) < \delta$

para cualquier par de números naturales n y m mayores o iguales a N.

TEOREMA 10.5. *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles que converge en medida, entonces $(f_n)_{n\in\mathbb{N}}$ es de Cauchy en medida.*

Demostración

Sea f el límite ne medida de la sucesión $(f_n)_{n\in\mathbb{N}}$.

Dadas $\varepsilon > 0$ y $\delta > 0$, sea $N \in \mathbb{N}$ tal que:

$\mu\left[|f_n. - f| > \frac{1}{2}\varepsilon\right] < \frac{1}{2}\delta$

para cualquier número natural $n \geq N$.

Entonces, si n y m son dos números naturales mayores o iguales que N, se tiene:

$\mu\left[|f_n. - f_m| > \varepsilon\right] \leq \mu\left[|f_n. - f| > \frac{1}{2}\varepsilon\right] + \mu\left[|f_m. - f| > \frac{1}{2}\varepsilon\right] < \delta.$

Por lo tanto, $(f_n)_{n\in\mathbb{N}}$ es de Cauchy en medida.

■

TEOREMA 10.6. *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles y supongamos que $(f_n)_{n\in\mathbb{N}}$ es de Cauchy en medida, entonces $(f_n)_{n\in\mathbb{N}}$ contiene una subsucesión $(f_{n_k})_{k\in\mathbb{N}}$ que converge casi en todas partes a una función medible f y tal que, dada $\delta > 0$, existe un conjunto medible A tal que $\mu(A) < \delta$ y la sucesión $(f_{n_k})_{k\in\mathbb{N}}$ converge uniformemente a f sobre A^c.*

Demostración

Para cada $k \in \mathbb{N}$, definamos $\varepsilon_k = \delta_k = \frac{1}{2^k}$. Sabemos que existe $N_k \in \mathbb{N}$ tal que:

$\mu\left(\{y \in \mathbb{F} : |f_n(y) - f_m(y)| > \varepsilon_k\}\right) < \delta_k$

para cualquier par de números naturales n y m mayores o iguales a N_k.

Existe entonces una sucesión creciente de números naturales $(m_k)_{k \in \mathbb{N}}$ tal que:

$\mu\left[\left|f_{m_{k+1}}. - f_{m_k}\right| > \frac{1}{2^k}\right] < \frac{1}{2^k}$ para cualquier $k \in \mathbb{N}$.

Sea $B = \bigcap_{j=1}^{\infty}\left(\bigcup_{i=j}^{\infty}\left[\left|f_{m_{i+1}}. - f_{m_i}\right| > \frac{1}{2^i}\right]\right)$, entonces, como:

$\sum_{k=1}^{\infty} \mu\left[\left|f_{m_{k+1}}. - f_{m_k}\right| > \frac{1}{2^k}\right] < \infty$,

por el lema de Borel-Cantelli, se tiene que $\mu(B) = 0$.

Para cada $j \in \mathbb{N}$, definamos $B_j = \bigcup_{i=j}^{\infty}\left[\left|f_{m_{i+1}}. - f_{m_i}\right| > \frac{1}{2^i}\right]$.

Para cualquier $i \in \mathbb{N}$ tal que $i \geq j$, se tiene:

$\left|f_{m_{i+1}}(y). - f_{m_i}(y)\right| \leq \frac{1}{2^i}$

para cualquier $y \in B_j^c$.

Dada $\varepsilon > 0$, sea $i_0 \in \mathbb{N}$ tal que $i_0 \geq N$ y $\sum_{i=i_0}^{\infty} \frac{1}{2^i} < \varepsilon$. Entonces, si $r > s \geq i_0$, se tiene:

$|f_{m_r}(y). - f_{m_s}(y)| \leq \sum_{i=s}^{r-1}\left|f_{m_{i+1}}(y). - f_{m_i}(y)\right|$

$\leq \sum_{i=s}^{r-1} \frac{1}{2^i} < \sum_{i=i_0}^{\infty} \frac{1}{2^i} < \varepsilon$

para cualquier $y \in B_j^c$.

Así que la sucesión $(f_{m_k})_{k \in \mathbb{N}}$ es uniformemente de Cauchy en B_j^c. Por lo tanto, converge uniformemente en B_j^c.

Si $y \in B^c = \bigcup_{j=1}^{\infty}\left(\bigcap_{i=j}^{\infty}\left[\left|f_{m_{i+1}}. - f_{m_i}\right| \leq \frac{1}{2^i}\right]\right)$, entonces existe $j \in \mathbb{N}$ tal que $y \in B_j^c$, así que la sucesión $(f_{m_k}(y))_{k \in \mathbb{N}}$ es convergente.

Sea $f(y) = \lim_{k \to \infty} f_{m_k}(y)$ .Entonces $f_{m_k} \xrightarrow{c.t.p.} f$ ya que $\mu(B) = 0$.

Además, para cualquier $j \in \mathbb{N}$, $(f_{m_k})_{k \in \mathbb{N}}$ converge uniformemente a f en B_j^c.

Finalmente, como $\lim_{j \rightsquigarrow \infty} \mu(B_j) = 0$, dada $\delta > 0$, existe $N \in \mathbb{N}$ tal que $\mu(B_N) < \delta$.

$\blacksquare$

COROLARIO 10.7. *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles que converge en medida, entonces $(f_n)_{n\in\mathbb{N}}$ contiene una subsucesión $(f_{n_k})_{k\in\mathbb{N}}$ que converge casi en todas partes a una función medible f y tal que, dada $\delta > 0$, existe un conjunto medible A tal que $\mu(A) < \delta$ y la sucesión $(f_{n_k})_{k\in\mathbb{N}}$ converge uniformemente a f sobre A^c.*

TEOREMA 10.7. *Supongamos que μ es finita y sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles, de Cauchy en medida, entonces $(f_n)_{n\in\mathbb{N}}$ converge en medida..*

Demostración

Sea $(f_{m_k})_{k\in\mathbb{N}}$ una subsucesión que converge casi en todas partes a la función f, entonces $f_{m_k} \xrightarrow{\mu} f$.

Dadas $\varepsilon > 0$ y $\delta > 0$, sea $N \in \mathbb{N}$ tal que:

$\mu\left[|f_n. - f_m| > \frac{1}{2}\varepsilon\right] < \frac{1}{2}\delta,$

$\mu\left[|f_{m_k}. - f| > \frac{1}{2}\varepsilon\right] < \frac{1}{2}\delta,$

para cualquier terna de números naturales, n, m y k, mayores o iguales que N.

Sean n y k números naturales mayores o iguales que N, entonces:

$\mu\left[|f_n. - f| > \varepsilon\right] \leq \mu\left[|f_n. - f_{m_k}| > \frac{1}{2}\varepsilon\right] + \mu\left[|f_{m_k}. - f| > \frac{1}{2}\varepsilon\right] < \delta.$

Por lo tanto $f_n \xrightarrow{\mu} f$.

■

Si μ es finita y $(f_n)_{n\in\mathbb{N}}$ es una sucesión de funciones medibles que converge casi en todas partes a la función medible f, entonces se tiene convergencia uniforme, fuera de un conjunto fijo de medida pequeña, para toda la sucesión $(f_n)_{n\in\mathbb{N}}$:

TEOREMA 10.8. *Si μ es finita y $(f_n)_{n\in\mathbb{N}}$ es una sucesión de funciones medibles que converge casi en todas a la función medible f, entonces, dada $\delta > 0$, existe un conjunto medible A tal que $\mu(A) < \delta$ y la sucesión $(f_n)_{n\in\mathbb{N}}$ converge uniformemente a f sobre A^c.*

Demostración

Para cualquier $\varepsilon > 0$, se tiene:

$\mu\left(\limsup\left\{y \in \mathbb{F} : |f_n(y) - f(y)| > \varepsilon\right\}\right) = 0.$

Para cada $i \in \mathbb{N}$, sea $B^{(i)} = \bigcap_{j=1}^{\infty}\left(\bigcup_{k=j}^{\infty}\left[|f_k. - f| > \frac{1}{2^i}\right]\right)$, entonces $\mu\left(B^{(i)}\right) = 0$.

Para cada $j \in \mathbb{N}$, definamos $B_j^{(i)} = \bigcup_{k=j}^{\infty}\left[|f_k. - f| > \frac{1}{2^i}\right]$.

Como $\lim_{j\rightsquigarrow\infty}\mu\left(B_j^{(i)}\right) = 0$, dada $\delta > 0$, existe $N_i \in \mathbb{N}$ tal que $\mu\left(B_{N_i}^{(i)}\right) < \frac{\delta}{2^i}$.

Definamos $B = \bigcup_{i=1}^{\infty} B_{N_i}^{(i)}$, entonces $\mu(B) < \delta$.

Para cualquier $k \in \mathbb{N}$ tal que $k \geq N_i$, se tiene:

$|f_k(y) . - f(y)| \leq \frac{1}{2^i}$

para cualquier $y \in \left(B_{N_i}^{(i)}\right)^c \supset B^c$.

Dada $\varepsilon > 0$, sea $i \in \mathbb{N}$ tal que $\frac{1}{2^i} < \varepsilon$. Entonces, para cualquier $k \geq N_i$, se tiene:

$|f_k(y) . - f(y)| < \varepsilon$

para cualquier $y \in B^c$.

Por lo tanto, $(f_n)_{n\in\mathbb{N}}$ converge uniformemente en B^c.

■

Teorema 10.9. *Sean g una función no negativa e integrable, f una función medible y $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tales que $|f_n| \leq g$ para cualquier $n \in \mathbb{N}$ y $f_n \xrightarrow{\mu} f$, entonces $\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|\, d\mu = 0$.*

Demostración

Para cada $n \in \mathbb{N}$, definamos:

$s_n = \int_{\mathbb{F}} |f_n - f|\, d\mu$.

Se tiene entonces:

$s_n = \int_{\mathbb{F}} |f_n - f|\, d\mu \leq \int_{\mathbb{F}} |f_n|\, d\mu + \int_{\mathbb{F}} |f|\, d\mu \leq 2\int_{\mathbb{F}} g d\mu < \infty$.

Así que la sucesión $(s_n)_{n\in\mathbb{N}}$ está acotada.

Sea $(s_{n_k})_{k\in\mathbb{N}}$ una subsucesión convergente. Como $f_n \xrightarrow{\mu} f$, existe una subsucesión $\left(s_{n_{k_j}}\right)_{j\in\mathbb{N}}$ de $(s_{n_k})_{k\in\mathbb{N}}$ tal que $f_{n_{k_j}} \xrightarrow{c.t.p.} f$. Por el teorema de la convergencia dominada, la sucesión $\left(s_{n_{k_j}}\right)_{j\in\mathbb{N}}$ converge a cero, así que $(s_{n_k})_{k\in\mathbb{N}}$ converge a cero. Por lo tanto, la sucesión $(s_n)_{n\in\mathbb{N}}$ converge a cero.

■

10.4. Convergencia débil

En esta sección asumiremos que la medida μ es finita.

Si $f : \mathbb{F} \to \mathbb{R}$ es una función medible denotaremos por ν_f la proyección de μ bajo f . En otras palabras, ν_f es la medida sobre los borelianos de $\mathbb{R}$ definida por:

$\nu_f(B) = \mu(\{y \in \mathbb{F} : f(y) \in B\})$.

Como sabemos, está medida queda únicamente determinada por sus valores en los intervalos de la forma $(-\infty, x]$, donde $x \in \mathbb{R}$.

Definición 10.4. *Diremos que una sucesión $(\nu_n)_{n\in\mathbb{N}}$ de medidas finitas, definidas sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, converge débilmente a la medida finita ν, definida sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, si se cumplen las siguientes dos condiciones:*

(i) $\lim_{n\rightsquigarrow\infty} \nu_n(\mathbb{R}) = \nu(\mathbb{R})$.
(ii) $\lim_{n\rightsquigarrow\infty} \nu_n((-\infty, x]) = \nu((-\infty, x])$ *para todo elemento* $x \in \mathbb{R}$ *tal que* $\nu(\{x\}) = 0$.

Si éste es el caso, se escribirá $\nu_n \xrightarrow{d} \nu$.

Obsérvese que se tiene:

$\nu(\{x\}) = \nu((-\infty, x]) - \nu((-\infty, x))$.

Así que, $\nu(\{x\}) = 0$ si y sólo si $\nu((-\infty, x)) = \nu((-\infty, x])$.

Además:

$\nu((-\infty, x)) = \lim_{\varepsilon\to 0+} \nu((-\infty, x-\varepsilon])$.

Así que, $\nu(\{x\}) = 0$ si y sólo si la función $x \to \nu((-\infty, x])$, definida sobre $\mathbb{R}$, es continua por la izquierda. Esta función siempre es continua por la derecha, así que:

$\nu(\{x\}) = 0$ si y sólo si la función $x \to \nu((-\infty, x])$, definida sobre $\mathbb{R}$, es continua.

Además, como ν es finita, el conjunto $\{x \in \mathbb{R} : \nu(\{x\}) > 0\}$ es a lo más infinito numerable.

Definición 10.5. *Si ν es una medida sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, diremos que $x \in \mathbb{R}$ es punto de continuidad de ν si $\nu(\{x\}) = 0$. ∞ y $-\infty$ serán considerados puntos de continuidad de P. Si I es un intervalo en $\mathbb{R}$, diremos que I es un intervalo de continuidad de ν si sus extremos son puntos de continuidad de ν.*

Definición 10.6. *Diremos que una función $g : \mathbb{R} \mapsto \mathbb{R}$ se anula en el infinito si dada cualquier $\varepsilon > 0$ existe un conjunto compacto $K \subset \mathbb{R}$ tal que $|g(x)| < \varepsilon$ para cualquier $x \notin K$.*

Definición 10.7. *Diremos que una función $g : \mathbb{R} \mapsto \mathbb{R}$ tiene soporte compacto si existe un conjunto compacto $K \subset \mathbb{R}$ tal que $g(x) = 0$ para cualquier $x \notin K$.*

Proposición 10.5. *Sea $(\nu_n)_{n\in\mathbb{N}}$ una sucesión de medidas finitas tales que $\nu_n \xrightarrow{d} \nu$ y $\nu_n \xrightarrow{d} \nu'$, entonces $\nu = \nu'$.*

Demostración

Si x es un número real tal que $\nu(\{x\}) = 0$ y $\nu'(\{x\}) = 0$, entonces $\nu((-\infty, x]) = \nu'((-\infty, x])$. Pero como los conjuntos $\{x \in \mathbb{R} : \nu(\{x\}) > 0\}$ y $\{x \in \mathbb{R} : \nu'(\{x\}) > 0\}$ son a lo más infinito numerables, el conjunto $\{x \in \mathbb{R} : \nu(\{x\}) = 0$ y $\nu'(\{x\}) =\}$ es denso en $\mathbb{R}$. Así que, siendo continuas por la derecha las funciones $x \to \nu((-\infty, x])$ y $x \to \nu'((-\infty, x])$, se sigue que $\nu((-\infty, x]) = \nu'((-\infty, x])$ para cualquier $x \in \mathbb{R}$. Por lo tanto $\nu(B) = \nu'(B)$ para cualquier conjunto boreliano de $\mathbb{R}$.

$\blacksquare$

Proposición 10.6. *Sea ν una medida finita sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ y $x \in \mathbb{R}$ un punto de continuidad de ν. Entonces, dada $\varepsilon > 0$, existe un intervalo de continuidad de ν, finito, (a, b), tal que $x \in (a, b)$ y $\nu[(a, b)] < \varepsilon$.*

Demostración

Como el conjunto $\{x \in \mathbb{R} : \nu(\{x\}) > 0\}$ es a lo más infinito numerable, el conjunto de puntos de continuidad de ν es denso en $\mathbb{R}$.

Sea $(a_n)_{n\in\mathbb{N}}$ una sucesión creciente de puntos de continuidad de ν tal que $\lim_{n\rightsquigarrow\infty} a_n = x$, y $(b_n)_{n\in\mathbb{N}}$ una sucesión decreciente de puntos de continuidad de ν tal que $\lim_{n\rightsquigarrow\infty} b_n = x$. Entonces:

$\lim_{n\rightsquigarrow\infty} \nu[(a_n, b_n)] = \nu[\bigcap_{n=1}^{\infty}(a_n, b_n)] = P(\{x\}) = 0,$

de lo cual se sigue el resultado.

$\blacksquare$

Teorema 10.10. *Si $(\nu_n)_{n\in\mathbb{N}}$ es una sucesión de medidas finitas sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ tales que $\lim_{n\rightsquigarrow\infty} \nu_n(\mathbb{R}) = \nu(\mathbb{R})$, donde ν es una medida finita sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, entonces las siguientes condiciones son equivalentes:*

(i) $\nu_n \xrightarrow{d} \nu$.
(ii) $\lim_{n\rightsquigarrow\infty} \nu_n(I) = \nu(I)$ *para todo intervalo de continuidad, I, de ν.*
(iii) $\lim_{n\rightsquigarrow\infty} \nu_n(I) = \nu(I)$ *para todo intervalo de continuidad, I, de ν, abierto y finito.*
(iv) $\lim_{n\rightsquigarrow\infty} \nu_n(I) = \nu(I)$ *para todo intervalo de continuidad, I, de ν, finito.*
(v) $\lim_{n\rightsquigarrow\infty} \int_{\mathbb{R}} g d\nu_n = \int_{\mathbb{R}} g d\nu$ *para cualquier función continua $g : \mathbb{R} \mapsto \mathbb{R}$ que se anula en el infinito.*
(vi) $\lim_{n\rightsquigarrow\infty} \int_{\mathbb{R}} g d\nu_n = \int_{\mathbb{R}} g d\nu$ *para cualquier función continua $g : \mathbb{R} \mapsto \mathbb{R}$ con soporte compacto.*
(vii) $\lim_{n\rightsquigarrow\infty} \int_{\mathbb{R}} g d\nu_n = \int_{\mathbb{R}} g d\nu$ *para cualquier función $g : \mathbb{R} \mapsto \mathbb{R}$ continua y acotada.*

Demostración

Vamos a demostrar primero que *i*, *ii* y *iii* son equivalentes.

$i \Rightarrow ii$

Sea $x \in \mathbb{R}$ un punto de continuidad de ν y, dada $\varepsilon > 0$, sea (a, b) un intervalo de continuidad de ν, finito, tal que $x \in (a, b)$ y $\nu[(a, b)] < \varepsilon$. Entonces:

$\limsup_{n\rightsquigarrow\infty} \nu_n(\{x\}) \leq \limsup_{n\rightsquigarrow\infty} \nu_n[(a, b]]$

$= \lim_{n \rightsquigarrow \infty} \nu_n [(-\infty, b]] - \lim_{n \rightsquigarrow \infty} \nu_n [(-\infty, a]]$

$= \nu [(-\infty, b]] - \nu [(-\infty, a]] = \nu [(a, b]] = \nu [(a, b)] < \varepsilon.$

Por lo tanto, $\lim_{n \rightsquigarrow \infty} \nu_n (\{x\}) = 0$.

Así que, $\lim_{n \rightsquigarrow \infty} \nu_n (I) = \nu (I)$ para todo intervalo de continuidad, I, de ν de la forma $I = (-\infty, x)$, donde $x \in \mathbb{R}$.

Además, como $\lim_{n \rightsquigarrow \infty} \nu_n (\mathbb{R}) = \nu (\mathbb{R})$, entonces $\lim_{n \rightsquigarrow \infty} \nu_n (I) = \nu (I)$ para todo intervalo de continuidad, I, de ν de la forma $I = (x, \infty)$ o $I = [x, \infty)$, donde $x \in \mathbb{R}$.

Finalmente, si I es un intervalo de continuidad de ν, finito, con extremos a y b, entonces:

$\lim_{n \rightsquigarrow \infty} \nu_n (I) = \lim_{n \rightsquigarrow \infty} \nu_n [(-\infty, b]] - \lim_{n \rightsquigarrow \infty} \nu_n [(-\infty, a]]$

$= \nu [(-\infty, b]] - \nu [(-\infty, a]] = \nu [(a, b]] = \nu (I).$

$ii \Rightarrow iii$ es inmediato ya que iii es un caso particular de ii.

$iii \Rightarrow i$

Sea C el conjunto de puntos de continuidad de ν, $x \in C$.

Dada $\varepsilon > 0$, sea $b \in C$ tal que $b > |x|$, $\nu ((-\infty, -b)) < \varepsilon$ y $\nu ((b, \infty)) < \varepsilon$, y sea $N \in \mathbb{N}$ tal que:

$|\nu_n ((-b, x)) - \nu ((-b, x))| < \varepsilon,$

$|\nu_n ((-b, b)) - \nu ((-b, b))| < \varepsilon,$

$|\nu_n (\mathbb{R}) - \nu (\mathbb{R})| < \varepsilon,$

para cualquier $n \geq N$.

Entonces, para $n \geq N$, se tiene:

$|\nu_n [(-\infty, x)] - \nu [(-\infty, x)]|$

$= |\nu_n [(-\infty, -b)] + \nu_n [(-b, x)] - \nu [(-\infty, -b)] - \nu [(-b, x)]|$

$\leq \nu_n [(-\infty, -b)] + \nu [(-\infty, -b)] + |\nu_n [(-b, x)] - \nu [(-b, x)]|$

$\leq \nu_n (\mathbb{R}) - \nu_n [(-b, b)] + \nu [(-\infty, -b)] + |\nu_n [(-b, x)] - \nu [(-b, x)]|$

$\leq |\nu_n (\mathbb{R}) - \nu (\mathbb{R})| + \nu (\mathbb{R}) - \nu_n [(-b, b)] + \nu [(-\infty, -b)] + |\nu_n [(-b, x)] - \nu [(-b, x)]|$

$= |\nu_n (\mathbb{R}) - \nu (\mathbb{R})| + \nu [(-\infty, -b)] + \nu [(-b, b)] + \nu [(b, \infty)]$

$+\nu [(-\infty, -b)] - \nu_n [(-b, b)] + |\nu_n [(-b, x)] - \nu [(-b, x)]|$

$\leq |\nu_n(\mathbb{R}) - \nu(\mathbb{R})| + |\nu_n[(-b,x)] - \nu[(-b,x)]|$

$+ |\nu_n[(-b,b)] - \nu[(-b,b)]| + 2\nu[(-\infty,-b)] + \nu[(b,\infty)]$

$< 6\varepsilon.$

Así que:

$\text{lím}_{n \rightsquigarrow \infty} \nu_n((-\infty,x)) = \nu((-\infty,x)).$

Queda entonces demostrado que i, ii y iii son equivalentes.

Ahora vamos a demostrar que iii, iv, v y vi son equivalentes.

$iii \Rightarrow iv$

Sea $x \in \mathbb{R}$ un punto de continuidad de ν.

Dada $\varepsilon > 0$, sea (a,b) un intervalo de continuidad de ν, finito, tal que $x \in (a,b)$ y $\nu[(a,b)] < \varepsilon$.

Se tiene:

$\text{lím}_{n \rightsquigarrow \infty} \nu_n((a,b)) = \nu((a,b)).$

Así que existe $N \in \mathbb{N}$ tal que:

$\nu_n(\{x\}) \leq \nu_n((a,b)) < \nu((a,b)) + \varepsilon < 2\varepsilon$

para cualquier $n \geq N$.

Por lo tanto, $\text{lím}_{n \rightsquigarrow \infty} \nu_n(\{x\}) = 0$.

Así que, si I es un intervalo de continuidad, I, de ν, finito, con extremos c y d, entonces:

$\text{lím}_{n \rightsquigarrow \infty} \nu_n(I) = \text{lím}_{n \rightsquigarrow \infty} \nu_n((c,d)) = \nu((c,d)) = \nu(I).$

$iv \Rightarrow v$

Sea $g : \mathbb{R} \mapsto \mathbb{R}$ una función continua que se anula en el infinito y, dada $\varepsilon > 0$, sea $[-a,a]$ un intervalo de continuidad de ν tal que $|g(x)| < \varepsilon$ para cualquier $x \notin [-a,a]$. Se tiene entonces:

$\int_{[-a,a]^c} |g|\, d\nu_n \leq \varepsilon \nu_n([-a,a]^c) \leq \varepsilon \nu(\mathbb{R}),$

$\int_{[-a,a]^c} |g|\, d\nu \leq \varepsilon \nu([-a,a]^c) \leq \varepsilon \nu(\mathbb{R}).$

Como g es uniformemente continua sobre el intervalo $[-a,a]$, existe una partición de $[-a,a]$ en intervalos ajenos, $I_1, \ldots, I_r$, todos intervalos de continuidad de ν, en cada uno de los cuales la oscilación de g es menor que ε.

Para cada $j \in \{1, \ldots, r\}$ tomemos $x_j \in I_j$, y definamos $T(x) = g(x_j)$ para cualquier $x \in I_j$. Se tiene entonces:

$\left|\int_{-a}^{a} gd\nu - \int_{-a}^{a} Td\nu\right| \leq \int_{-a}^{a} |g - T|\, d\nu \leq \varepsilon\nu([-a, a]) \leq \varepsilon\nu(\mathbb{R})$,

$\left|\int_{-a}^{a} gd\nu_n - \int_{-a}^{a} Td\nu_n\right| \leq \int_{-a}^{a} |g - T|\, d\nu_n \leq \varepsilon\mu_n([-a, a]) \leq \varepsilon\nu_n(\mathbb{R})$.

Por otra parte, como $\lim_{n\rightsquigarrow\infty} \nu_n(I_j) = \nu(I_j)$ para cualquier $j \in \{1, \ldots, r\}$, se tiene:

$\lim_{n\rightsquigarrow\infty} \int_{-a}^{a} Td\nu_n = \lim_{n\rightsquigarrow\infty} \sum_{j=1}^{n} x_j\nu_n(I_j) = \sum_{j=1}^{n} x_j\mu(I_j) = \int_{-a}^{a} Td\nu$.

Así que, existe $N \in \mathbb{N}$ tal que $\left|\int_{-a}^{a} Td\mu_n - \int_{-a}^{a} Td\nu\right| < \varepsilon$ para cualquier $n \geq N$.

Por lo tanto, para cualquier $n \geq N$, se tiene:

$\left|\int_{\mathbb{R}} gd\nu_n - \int_{\mathbb{R}} gd\nu\right| = \left|\int_{-a}^{a} gd\nu_n + \int_{[-a,a]^c} gd\mu_n - \int_{-a}^{a} gd\nu - \int_{[-a,a]^c} gd\nu\right|$

$\leq \int_{[-a,a]^c} |g|\, d\nu_n + \int_{[-a,a]^c} |g|\, d\nu + \left|\int_{-a}^{a} gd\nu_n - \int_{-a}^{a} Td\nu_n\right|$

$+ \left|\int_{-a}^{a} Td\nu_n - \int_{-a}^{a} Td\nu\right| + \left|\int_{-a}^{a} gd\nu - \int_{-a}^{a} Td\nu\right| < 5\varepsilon(\nu(\mathbb{R}) + 1)$.

Así que, $\lim_{n\rightsquigarrow\infty} \left|\int_{\mathbb{R}} gd\mu_n - \int_{\mathbb{R}} gd\nu\right| = 0$.

$v \Rightarrow vi$ es inmediato ya que toda función continua con soporte compacto se anula en el infinito.

$vi \Rightarrow iii$

Sea (a, b) un intervalo de continuidad de ν, abierto y finito.

Dada $\varepsilon > 0$, sea $\delta > 0$ tal que $\delta < \frac{1}{2}(b - a)$, $a - \delta$, $a + \delta$, $b - \delta$ y $b + \delta$ son puntos de continuidad de ν y $\nu[(a - \delta, a + \delta)] + \nu[(b - \delta, b + \delta)] < \varepsilon$.

Sea $g_1 : \mathbb{R} \mapsto \mathbb{R}$ la función definida como sigue:

$$g_1(x) = \begin{cases} 1 & \text{si } x \in [a, b] \\ 1 - \frac{1}{\delta}(a - x) & \text{si } x \in (a - \delta, a) \\ 1 - \frac{1}{\delta}(x - b) & \text{si } x \in (b, b + \delta) \\ 0 & \text{en otro caso} \end{cases}$$

g_1 es no negativa, continua, tiene soporte compacto y está acotada por 1.

Se tiene entonces:

$\lim_{n\rightsquigarrow\infty} \int_{\mathbb{R}} g_1 d\nu_n = \int_{\mathbb{R}} g_1 d\nu$.

Así que, dada $\varepsilon > 0$, existe $N_1 \in \mathbb{N}$ tal que:

$\int_{\mathbb{R}} g_1 d\nu_n < \int_{\mathbb{R}} g_1 d\nu + \varepsilon$

para cualquier $n \geq N_1$.

Además, se tiene:

$\int_{\mathbb{R}} g_1 d\nu_n = \int_{a-\delta}^{b+\delta} g_1 d\mu_n \geq \int_a^b g_1 d\nu_n = \nu_n [(a,b)]$,

$\int_{\mathbb{R}} g_1 d\nu = \int_{a-\delta}^{b+\delta} g_1 d\nu \leq \nu [(a-\delta, b+\delta)]$

$= \nu [(a,b)] + \nu [(a-\delta, a+\delta)] + \nu [(b-\delta, b+\delta)] < \nu [(a,b)] + \varepsilon$.

Así que, para $n \geq N_1$, se tiene:

$\nu_n [(a,b)] \leq \int_{\mathbb{R}} g_1 d\mu_n < \int_{\mathbb{R}} g_1 d\nu + \varepsilon$

$\leq \nu [(a-\delta, b+\delta)] + \varepsilon < \nu [(a,b)] + 2\varepsilon$.

Definamos ahora $g_2 : \mathbb{R} \mapsto \mathbb{R}$ de la siguiente manera:

$$g_2(x) = \begin{cases} 1 & \text{si } x \in [a+\delta, b-\delta] \\ \frac{1}{\delta}(x-a) & \text{si } x \in (a, a+\delta) \\ \frac{1}{\delta}(b-x) & \text{si } x \in (b-\delta, b) \\ 0 & \text{en otro caso} \end{cases}$$

g_2 es no negativa, continua, tiene soporte compacto y está acotada por 1.

Se tiene entonces:

$\lim_{n \rightsquigarrow \infty} \int_{\mathbb{R}} g_2 d\nu_n = \int_{\mathbb{R}} g_1 d\nu$.

Así que, dada $\varepsilon > 0$, existe $N_2 \in \mathbb{N}$ tal que:

$\int_{\mathbb{R}} g_2 d\nu_n > \int_{\mathbb{R}} g_2 d\nu - \varepsilon$

para cualquier $n \geq N_2$.

Además, se tiene:

$\int_{\mathbb{R}} g_2 d\nu_n = \int_a^b g_2 d\nu_n \leq \nu_n [(a,b)]$

$\int_{\mathbb{R}} g_2 d\nu = \int_a^b g_2 d\nu \geq \int_{a+\delta}^{b-\delta} g_2 d\nu = \nu [(a+\delta, b-\delta)]$

$= \nu [(a,b)] - \nu [(a-\delta, a+\delta)] - \nu [(b-\delta, b+\delta)] > \nu [(a,b)] - \varepsilon$.

Así que, para $n \geq N_2$, se tiene:

$\nu_n [(a,b)] \geq \int_{\mathbb{R}} g_2 d\mu_n > \int_{\mathbb{R}} g_2 d\nu - \varepsilon$

$\geq \nu [(a+\delta, b-\delta)] - \varepsilon > \nu [(a,b)] - 2\varepsilon.$

Por lo tanto, para $n \geq \text{máx} \{N_1, N_2\}$, se tiene:

$\nu [(a,b)] - 2\varepsilon < \nu_n [(a,b)] < \nu [(a,b)] + 2\varepsilon.$

Así que, $\text{lím}_{n \rightsquigarrow \infty} \nu_n [(a,b)] = \nu [(a,b)]$.

Queda entonces demostrado que *iii*, *iv*, *v* y *vi* son equivalentes.

Siendo equivalentes *i*, *ii* y *iii*, hasta aquí tenemos demostrado que *i*, *ii*, *iii*, *iv*, *v* y *vi* son equivalentes.

Para completar la demostración vamos a probar que *vii* se sigue de *ii* y que *vi* se sigue de *vii*.

$ii \Rightarrow vii$

Sea $g : \mathbb{R} \mapsto \mathbb{R}$ una función continua y acotada y M una cota de $|g|$.

Como el conjunto de puntos de continuidad de ν es denso en $\mathbb{R}$ y:

$\text{lím}_{x \rightsquigarrow \infty} \nu [(-\infty, -x)] = \text{lím}_{x \rightsquigarrow \infty} \nu [(x, \infty)] = 0,$

dada $\varepsilon > 0$, existe $a > 0$, punto de continuidad de ν, tal que:

$\nu [(-\infty, -a]] + \nu [[a, \infty)] < \varepsilon.$

Como se tiene:

$\text{lím}_{n \rightsquigarrow \infty} \nu_n [(-\infty, -a]] = \nu [(-\infty, -a]],$

$\text{lím}_{n \rightsquigarrow \infty} \nu_n [[a, \infty)] = \nu [[a, \infty)],$

existe $N_1 \in \mathbb{N}$ tal que:

$\nu_n [(-\infty, -a]] + \nu_n [[a, \infty)] < \nu [(-\infty, -a]] + \nu [[a, \infty)] + \varepsilon < 2\varepsilon$

para cualquier $n \geq N_1$.

También, como se tiene:

$\text{lím}_{n \rightsquigarrow \infty} \nu_n (\mathbb{R}) = \nu (\mathbb{R}),$

existe una constante K tal que $\nu (\mathbb{R}) < K$ y $\mu_n (\mathbb{R}) < K$ para cualquier $n \in \mathbb{N}$.

Como g es uniformemente continua sobre el intervalo $[-a,a]$, existe una partición de $[-a,a]$ en intervalos ajenos, $I_1,\ldots,I_r$, todos intervalos de continuidad de ν, en cada uno de los cuales la oscilación de g es menor que ε.

Para cada $j \in \{1,\ldots,r\}$ tomemos $x_j \in I_j$, y definamos $T(x) = g(x_j)$ para cualquier $x \in I_j$.

Se tiene entonces, para cualquier $n \geq N_1$:

$\left|\int_{-a}^{a} g d\nu - \int_{-a}^{a} T d\nu\right| \leq \int_{-a}^{a} |g-T|\, d\nu < \varepsilon\nu([-a,a]) \leq \varepsilon\nu(\mathbb{R}) < \varepsilon K,$

$\left|\int_{-a}^{a} g d\nu_n - \int_{-a}^{a} T d\nu_n\right| \leq \int_{-a}^{a} |g-T|\, d\nu_n < \varepsilon\nu_n([-a,a])$

$\leq \varepsilon\nu_n(\mathbb{R}) < \varepsilon K,$

$\left|\int_{-\infty}^{-a} g d\nu + \int_{a}^{\infty} g d\nu\right| \leq \int_{-\infty}^{-a} |g|\, d\nu + \int_{a}^{\infty} |g|\, d\nu$

$\leq M(\nu[(-\infty,-a]] + \nu[[a,\infty)]) < \varepsilon M,$

$\left|\int_{-\infty}^{-a} g d\nu_n + \int_{a}^{\infty} g d\mu_n\right| \leq \int_{-\infty}^{-a} |g|\, d\mu_n + \int_{a}^{\infty} |g|\, d\nu_n$

$\leq M(\nu_n[(-\infty,-a]] + \nu_n[[a,\infty)]) < 2\varepsilon M,$

Por otra parte, como $\lim_{n\rightsquigarrow\infty} \nu_n(I_j) = \nu(I_j)$ para cualquier $j \in \{1,\ldots,r\}$, se tiene:

$\lim_{n\rightsquigarrow\infty} \int_{-a}^{a} T d\nu_n = \lim_{n\rightsquigarrow\infty} \sum_{j=1}^{n} x_j\nu_n(I_j) = \sum_{j=1}^{n} x_j\mu(I_j) = \int_{-a}^{a} T d\nu.$

Así que, existe $N_2 \in \mathbb{N}$ tal que $\left|\int_{-a}^{a} T d\mu_n - \int_{-a}^{a} T d\nu\right| < \varepsilon$ para cualquier $n \geq N_2$.

Por lo tanto, para cualquier $n \geq \max\{N_1, N_2\}$, se tiene:

$\left|\int_{\mathbb{R}} g d\nu_n - \int_{\mathbb{R}} g d\nu\right|$

$\leq \left|\int_{-a}^{a} g d\nu_n - \int_{-a}^{a} g d\nu\right| + \left|\int_{-\infty}^{-a} g d\nu_n + \int_{a}^{\infty} g d\mu_n\right| + \left|\int_{-\infty}^{-a} g d\nu + \int_{a}^{\infty} g d\nu\right|$

$\leq \left|\int_{-a}^{a} g d\nu_n - \int_{-a}^{a} T d\nu_n\right| + \left|\int_{-a}^{a} T d\nu_n - \int_{-a}^{a} T d\nu\right|$

$+ \left|\int_{-a}^{a} g d\nu - \int_{-a}^{a} T d\nu\right| + 3\varepsilon M < 3\varepsilon + 3\varepsilon M = 3(M+1)\varepsilon.$

Así que, $\lim_{n\rightsquigarrow\infty} \left|\int_{\mathbb{R}} g d\mu_n - \int_{\mathbb{R}} g d\nu\right| = 0.$

vii $\Rightarrow$ *vi* es inmediato ya que toda función continua con soporte compacto es acotada.

$\blacksquare$

10.5. Convergencia en distribución

En esta sección asumiremos que la medida μ es finita.

Definición 10.8. *Sean $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles $f_n:\mathbb{F}\to\mathbb{R}$ y $f:\mathbb{F}\to\mathbb{R}$ una función medible . Diremos que la sucesión $(f_n)_{n\in\mathbb{N}}$ converge en distribución a f si $\nu_{f_n}\xrightarrow{d}\nu_f$. Se éste es el caso, escribiremos $f_n\xrightarrow{D}f$.*

Teorema 10.11. *Sean $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles y $f:\mathbb{F}\to\mathbb{R}$ una función medible tales que $f_n\xrightarrow{\mu}f$, entonces $f_n\xrightarrow{D}f$.*

Demostración

Se tiene $\nu(\mathbb{R})=\mu(\mathbb{F})$ y $\nu_n(\mathbb{R})=\mu(\mathbb{F})$, así que $\lim_{n\rightsquigarrow\infty}\nu_n(\mathbb{R})=\nu(\mathbb{R})$.

Para $\varepsilon>0$, $n\in\mathbb{N}$ y $t\in\mathbb{R}$, se tiene:

$\nu((-\infty,t-\varepsilon])=\mu\{y\in\mathbb{F}:f(y)\le t-\varepsilon\}$

$=\mu\{y\in\mathbb{F}:f(y)\le t-\varepsilon \text{ y } |f(y)-f_n(y)|>\varepsilon\}$

$+\mu\{y\in\mathbb{F}:f(y)\le t-\varepsilon \text{ y } |f(y)-f_n(y)|\le\varepsilon\}$

$=\mu\{y\in\mathbb{F}:f(y)\le t-\varepsilon \text{ y } |f(y)-f_n(y)|>\varepsilon\}$

$+\mu\{y\in\mathbb{F}:f(y)\le t-\varepsilon \text{ y } f(y)-\varepsilon\le f_n(y)\le f(y)+\varepsilon\}$

$\le\mu\{y\in\mathbb{F}:|f(y)-f_n(y)|>\varepsilon\}+\mu\{y\in\mathbb{F}:f_n(y)\le t\}$

$=\mu\{y\in\mathbb{F}:|f(y)-f_n(y)|>\varepsilon\}+\nu_n((-\infty,t]))$.

Así que:

$\nu((-\infty,t-\varepsilon])-\mu\{y\in\mathbb{F}:|f(y)-f_n(y)|>\varepsilon\}\le\nu_n((-\infty,t]))$.

Además:

$\nu_n((-\infty,t]))=\mu\{y\in\mathbb{F}:f_n(y)\le t\}$

$=\mu\{y\in\mathbb{F}:f(y)\le t \text{ y } |f(y)-f_n(y)|>\varepsilon\}$

$+\mu\{y\in\mathbb{F}:f(y)\le t \text{ y } |f(y)-f_n(y)|\le\varepsilon\}$

$=\mu\{y\in\mathbb{F}:f(y)\le t \text{ y } |f(y)-f_n(y)|>\varepsilon\}$

$+\mu\{y\in\mathbb{F}:f(y)\le t \text{ y } f_n(y)-\varepsilon\le f(y)\le f_n(y)+\varepsilon\}$

$\le\mu\{y\in\mathbb{F}:|f(y)-f_n(y)|>\varepsilon\}+\mu\{y\in\mathbb{F}:f(y)\le t+\varepsilon\}$

$=\mu\{y\in\mathbb{F}:|f(y)-f_n(y)|>\varepsilon\}+\nu((-\infty,t+\varepsilon]))$.

Por lo tanto:

$\nu\left((-\infty, t-\varepsilon]\right) - \mu\left\{y \in \mathbb{F} : |f(y) - f_n(y)| > \varepsilon\right\} \leq \nu_n\left((-\infty, t]\right)$

$\leq \mu\left\{y \in \mathbb{F} : |f(y) - f_n(y)| > \varepsilon\right\} + \nu\left((-\infty, t+\varepsilon]\right).$

Tomando límites cuando $n \rightsquigarrow \infty$ y utilizando el hecho de que $f_n \xrightarrow{\mu} f$, se obtiene:

$\nu\left((-\infty, t-\varepsilon]\right) \leq \liminf_{n\rightsquigarrow\infty} \nu_n\left((-\infty, t]\right) \leq \limsup_{n\rightsquigarrow\infty} \nu_n\left((-\infty, t]\right) \leq \nu\left((-\infty, t+\varepsilon]\right).$

Ahora, si t es un punto tal que $\nu(\{t\}) = 0$, entonces, tomando límites cuando $\varepsilon \to 0$, se obtiene:

$\nu\left((-\infty, t]\right) \leq \liminf_{n\to\infty} \nu_n\left((-\infty, t]\right) \leq \limsup_{n\rightsquigarrow\infty} \nu_n\left((-\infty, t]\right) \leq \nu\left((-\infty, t]\right).$

Así que $\lim_{n\rightsquigarrow\infty} \nu_n\left((-\infty, t]\right) = \nu\left((-\infty, t]\right)$, es decir, $\nu_n \xrightarrow{d} \nu$.

∎

Como se muestra en el siguiente ejemplo, el inverso del resultado anterior no es válido en general.

Ejemplo 10.3. *Sea $\mathbb{F} = (0,1]$ y μ la medida de Lebesgue sobre $\mathbb{F}$. Para cada $n \in \{0, 1, 2, \ldots\}$, definamos $f_n : \mathbb{F} \to \mathbb{R}$ de la siguiente manera:*

$$f_n(y) = \begin{cases} 2y - 1 & si\ n = 0\ o\ si\ n\ es\ impar \\ 1 - 2y & si\ n\ es\ par \end{cases}$$

Para $n \in \{0, 1, 3, 5, \ldots\}$ y $t \in \mathbb{R}$, se tiene:

$$\mu\left\{y \in \mathbb{F} : f_n(y) \leq t\right\} = \mu\left\{y \in \mathbb{F} : y \leq \tfrac{1}{2}(1+t)\right\} = \begin{cases} 0 & si\ t < -1 \\ \frac{1}{2}(1+t) & si\ t \in [-1, 1) \\ 1 & si\ t \geq 1 \end{cases}$$

Para $n \in \{2, 4, 6, \ldots\}$ y $t \in \mathbb{R}$, se tiene:

$$\mu\left\{y \in \mathbb{F} : f_n(y) \leq t\right\} = \mu\left\{y \in \mathbb{F} : y \geq \tfrac{1}{2}(1-t)\right\} = \begin{cases} 0 & si\ t < -1 \\ 1 - \frac{1}{2}(1-t) & si\ t \in [-1, 1) \\ 1 & si\ t \geq 1 \end{cases}$$

$$= \begin{cases} 0 & si\ t < -1 \\ \frac{1}{2}(1+t) & si\ t \in [-1, 1) \\ 1 & si\ t \geq 1 \end{cases}$$

Así que:

$$\nu_{f_0}\left((-\infty, t]\right) = \begin{cases} 0 & si\ t < -1 \\ \frac{1}{2}(1+t) & si\ t \in [-1, 1) \\ 1 & si\ t \geq 1 \end{cases}$$

$$\nu_{f_n}((-\infty,t]) = \begin{cases} 0 & si\ t < -1 \\ \frac{1}{2}(1+t) & si\ t \in [-1,1) \\ 1 & si\ t \geq 1 \end{cases}$$

Por lo tanto, $\nu_{f_n} \xrightarrow{d} \nu_{f_0}$.

Por otra parte, $|f_{2n} - f_0| = 2|f_0|$ *para cualquier* $n \in \mathbb{N}$, *así que, dada* $\varepsilon \in (0,1)$, *se tiene:*

$\mu\{y \in \mathbb{F} : |f_{2n}(y) - f_0(y)| > \varepsilon\} = \mu\left\{y \in \mathbb{F} : |f_0(y)| > \frac{\varepsilon}{2}\right\}$

$= \mu\left\{y \in \mathbb{F} : y > \frac{1}{2} + \frac{\varepsilon}{4}\right\} + \mu\left\{y \in \mathbb{F} : y < \frac{1}{2} - \frac{\varepsilon}{4}\right\} = 1 - \frac{\varepsilon}{2}.$

Por lo tanto:

$\lim_{n\rightsquigarrow\infty} \mu\{y \in \mathbb{F} : |f_{2n} - f_0| > \varepsilon\} = 1 - \frac{\varepsilon}{2} > 0.$

Así que la sucesión $(f_n)_{n\in\mathbb{N}}$ *no converge a* f_0 *en medida.*

Además, para cualquier $n \in \mathbb{N}$, $f_{2n-1}(y) = f_0(y)$ *para cualquier* $y \in \mathbb{F}$. *Por lo tanto, la sucesión* $(f_{2n-1})_{n\in\mathbb{N}}$ *converge en medida a* f_0.

Ahora bien, si la sucesión $(f_n)_{n\in\mathbb{N}}$ *convergiera en medida a una función medible* $g : \mathbb{F} \to \mathbb{R}$, *entonces la sucesión* $(f_{2n-1})_{n\in\mathbb{N}}$ *convergería en medida a* g. *Así que se tendría* $f_0 = g$ *casi en todas partes. Pero entonces se tendría que la sucesión* $(f_n)_{n\in\mathbb{N}}$ *converge en medida a* f_0, *lo cual es falso.*

Podemos concluir entonces que la sucesión $(f_n)_{n\in\mathbb{N}}$ *no converge en medida.*

Se tiene el siguiente resultado parcial:

TEOREMA 10.12. *Sean* $(f_n)_{n\in\mathbb{N}}$ *una sucesión de funciones medibles* $f_n : \mathbb{F} \to \mathbb{R}$. *Supongamos que* $\nu_{f_n} \xrightarrow{d} \nu_0$, *donde* ν_0 *es una medida finita concentrada en cero. Entonces* $f_n \xrightarrow{\mu} 0$.

Demostración

La hipótesis nos dice que:

$$\lim_{n\rightsquigarrow\infty} \nu_{f_n}((-\infty,x]) = \begin{cases} 0 & \text{si } x < 0 \\ \nu_0(\{0\}) & \text{si } x > 0 \end{cases}$$

Además, para cualquier $\varepsilon > 0$, se tiene:

$\mu(\{y \in \mathbb{F} : |f_n(y)| > \varepsilon\}) = \mu(\{y \in \mathbb{F} : f_n(y) > \varepsilon\}) + \mu(\{y \in \mathbb{F} : f_n(y) < -\varepsilon\})$

$\leq \mu(\{y \in \mathbb{F} : f_n(y) > \varepsilon\}) + \mu(\{y \in \mathbb{F} : f_n(y) \leq -\varepsilon\})$

$= \nu_{f_n}(\mathbb{R}) - \nu_{f_n}((-\infty,\varepsilon]) + \nu_{f_n}((-\infty,-\varepsilon]).$

Así que, $\limsup_{n\rightsquigarrow\infty} \mu(\{y \in \mathbb{F} : |f_n(y)| > \varepsilon\})$

$\leq \operatorname{lím}_{n\rightsquigarrow\infty} \nu_{f_n}(\mathbb{R}) - \operatorname{lím}_{n\rightsquigarrow\infty} \nu_{f_n}((-\infty,\varepsilon]) + \operatorname{lím}_{n\rightsquigarrow\infty} \nu_{f_n}((-\infty,-\varepsilon])$

$= \nu_0(\mathbb{R}) - \nu_0(\{0\}) = 0.$

Por lo tanto, $f_n \xrightarrow{\mu} 0$.

■

Corolario 10.8. *Sean $(f_n)_{n\in\mathbb{N}}$ y $(f'_n)_{n\in\mathbb{N}}$ dos sucesiones de funciones medibles. Supongamos que $f_n \xrightarrow{D} 0$ y $f'_n \xrightarrow{D} 0$, entonces $f_n + f'_n \xrightarrow{D} 0$ y $.f_n f'_n \xrightarrow{D} 0$.*

Demostración

Como $f_n \xrightarrow{D} 0$ y $f'_n \xrightarrow{D} 0$, se tiene que $\nu_{f_n} \xrightarrow{d} \nu_0$ y $\nu_{f'_n} \xrightarrow{d} \nu_0$, donde ν_0 es una medida finita concentrada en cero. Así que, $f_n \xrightarrow{\mu} 0$ y $f'_n \xrightarrow{\mu} 0$. Por lo tanto, $f_n + f'_n \xrightarrow{\mu} 0$ y $.f_n f'_n \xrightarrow{\mu} 0$, lo cual implica que $f_n + f'_n \xrightarrow{D} 0$ y $.f_n f'_n \xrightarrow{D} 0$.

■

CAPÍTULO 11

ESPACIOS L^p

11.1. Espacios L^p

Para $p \in (0, \infty)$, denotaremos por $\mathcal{L}^p$ al conjunto de funciones medibles f tales que $\int_{\mathbb{F}} |f|^p \, d\mu < \infty$. También, denotaremos por $\mathcal{L}^\infty$ al conjunto de funciones medibles y acotadas excepto a lo más en un conjunto de medida cero.

Obsérvese que si $f \in \mathcal{L}^p$, entonces f es finita casi en todas partes.

Para $p \in (0, \infty]$, el conjunto de clases de equivalencia en las cuales queda partido $\mathcal{L}^p$, mediante la relación de equivalencia definida por la igualdad casi en todas partes, será denotado por L^p.

Cada elemento de L^p es un conjunto de funciones con la propiedad de que cualquier par de ellas son iguales casi en todas partes.

Si $f : \mathbb{F} \to \overline{\mathbb{R}}$, la notación $f \in L^p$ significará que la clase de equivalencia de la cual forma parte f, pertenece a L^p, de manera que cualquier propiedad que se demuestre para f será en realidad una propiedad de la clase de equivalencia de la cual forma parte.

Si $f \in L^\infty$, diremos que $M \in \mathbb{R}$ es cota esencial de f si $|f| \leq M$ casi en todas partes. Además, definimos el supremo esencial de f, $\sup es\,(f)$, de la siguiente manera:

$\sup es\,(f) = \inf \{M \in \mathbb{R} : M \text{ es cota esencial de } f\}$

Obsérvese que si $f \in L^\infty$, entonces $|f| \leq \sup es\,(f)$ casi en todas partes. Además, no hay ningún número real M menor que $\sup es\,(f)$ y tal que $|f| \leq M$ casi en todas partes. Es decir, si $|f| \leq M$ casi en todas partes, entonces $\sup es\,(f) \leq M$.

Si $p \in [1, \infty)$ y $f \in L^p$, definimos $\|f\|_p = \left(\int_{\mathbb{F}} |f|^p \, d\mu \right)^{\frac{1}{p}}$.

Si $f \in L^\infty$, definimos $\|f\|_\infty = \sup es\,(|f|)$.

Obsérvese que se tienen las siguientes relaciones:

Si $p \in (0,1]$, entonces $y < y^p$ si $y \in (0,1)$ y $y > y^p$ si $y \in (1,\infty)$.

Si $p \in (1,\infty)$, entonces $y > y^p$ si $y \in (0,1)$ y $y < y^p$ si $y \in (1,\infty)$.

Para cualquier $p \in (0,\infty)$, la función $y \mapsto y^p$ es creciente en el intervalo $[0,\infty)$.

Si $r > p > 0$, entonces $y^r < y^p$ para cualquier $y \in (0,1)$ y $y^r > y^p$ para cualquier $y \in (1,\infty)$. También $y^p < 1 + y^r$ para cualquier $y \in [0,\infty)$.

Lema 11.1. *Sean $a, b \in \mathbb{R}$ y $p \in [0,\infty)$, entonces:*

$|a+b|^p \leq 2^p \left(|a|^p + |b|^p\right)$.

Demostración

$|a+b|^p \leq (|a| + |b|)^p \leq (2 \max\{|a|, |b|\})^p = 2^p \max\{|a|^p, |b|^p\} \leq 2^p \left(|a|^p + |b|^p\right)$.

■

Proposición 11.1. *Para cualquier $p \in (0,\infty]$, L^p es un espacio vectorial sobre $\mathbb{R}$.*

Demostración

Si $c \in \mathbb{R}$, es inmediato que si $f \in L^p$, entonces $cf \in L^p$.

Supongamos que $f, g \in L^p$, entonces, si $p \in [0,\infty)$, por el lema anterior, se tiene:

$\int_{\mathbb{F}} |f+g|^p \, d\mu \leq 2^p \left(\int_{\mathbb{F}} |f|^p \, d\mu + \int_{\mathbb{F}} |g|^p \, d\mu \right) < \infty$.

Así que $f + g \in L^p$.

Si $f, g \in L^\infty$, entonces, f y g son medibles y acotadas excepto a lo más en un conjunto de medida cero; por lo tanto, $f + g$ tiene la misma propiedad, así que $f + g \in L^\infty$.

Por lo tanto, en cualquier caso, L^p es un espacio vectorial sobre $\mathbb{R}$.

■

Lema 11.2. *Sea $t \in (0,1) \cup (1,\infty)$ y $\lambda \in (0,1)$, entonces:*

$t^\lambda < \lambda t + (1-\lambda)$.

Demostración

Para $t \in (0,\infty)$, definamos $f(t) = \lambda t + (1-\lambda) - t^\lambda$.

Se tiene $f'(t) = \lambda \left(1 - t^{\lambda-1}\right)$ para cualquier $t \in (0,\infty)$. Así que $f'(t) \leq 0$ para $t \in (0,1]$ y $f'(t) \geq 0$ para $t \in [1,\infty)$. Por lo tanto, f es decreciente en el intervalo $(0,1]$ y creciente en el intervalo $[1,\infty)$. Se concluye entonces que, para cualquier $t \in (0,1) \cup (1,\infty)$, $f(t) > f(1) = 0$, es decir, $\lambda t + (1-\lambda) > t^\lambda$.

■

COROLARIO 11.1. *Sean* $p, q \in (1, \infty)$ *tales que* $\frac{1}{p} + \frac{1}{q} = 1$ *y* $\alpha, \beta \in [0, \infty)$. *Entonces:*

$\alpha\beta \leq \frac{1}{p}\alpha^p + \frac{1}{q}\beta^q$.

La iguadad se cumple si y sólo si $\alpha^p = \beta^q$.

Demostración

Aplicando el lema anterior con $t = \frac{\alpha^p}{\beta^q}$ y $\lambda = \frac{1}{p}$, se tiene:

$\frac{\alpha}{\beta^{\frac{q}{p}}} < \frac{1}{p}\frac{\alpha^p}{\beta^q} + \frac{1}{q}$.

Así que:

$\alpha < \frac{1}{p}\frac{\alpha^p}{\beta} + \frac{1}{q}\beta^{\frac{q}{p}} = \alpha < \frac{1}{p}\frac{\alpha^p}{\beta} + \frac{1}{q}\beta^{q\left(1-\frac{1}{q}\right)}$.

Por lo tanto:

$\alpha\beta < \frac{1}{p}\alpha^p + \frac{1}{q}\beta^q$.

∎

TEOREMA 11.1 (**Desigualdad de Hölder**). *Sean* $p, q \in [1, \infty]$ *tales que* $\frac{1}{p} + \frac{1}{q} = 1$, $f \in L^p$ *y* $g \in L^q$, *entonces* $fg \in L^1$ *y:*

$||fg||_1 \leq ||f||_p \, ||g||_q$.

Demostración

Si $p, q \in (1, \infty)$, definamos $\alpha = \frac{|f|}{||f||_p}$ y $\beta = \frac{|g|}{||g||_q}$. Entonces:

$\left|\frac{f}{||f||_p}\frac{g}{||g||_q}\right| = \alpha\beta \leq \frac{1}{p}\alpha^p + \frac{1}{q}\beta^q = \frac{1}{p}\frac{|f|^p}{\left(||f||_p\right)^p} + \frac{1}{q}\frac{|g|^q}{\left(||g||_q\right)^q}$.

Integrando, se obtiene:

$\frac{1}{||f||_p||g||_q}\int_{\mathbb{F}} |fg| \, d\mu \leq \frac{1}{p} + \frac{1}{q} = 1$.

Si $f \in L^1$ y $g \in L^\infty$, entonces:

$\int_{\mathbb{F}} |fg| \, d\mu \leq \sup es\,(f) \int_{\mathbb{F}} |f| \, d\mu = ||f||_1 \, ||g||_\infty$.

∎

TEOREMA 11.2 (**Desigualdad de Minkowski**). *Sea* $p \in [1, \infty]$ *y* $f, g \in L^p$, *entonces:*

$||f + g||_p \leq ||f||_p + ||g||_p$.

Demostración

Para $p = 1$ se tiene:

$|f+g| \leq |f| + |g|$.

Así que:

$\int_{\mathbb{F}} |f+g| \, d\mu \leq \int_{\mathbb{F}} |f| \, d\mu + \int_{\mathbb{F}} |g| \, d\mu$.

Si $p \in (1, \infty)$, se tiene:

$\left[(f+g)^{p-1}\right]^q = (f+g)^{pq-q} = (f+g)^p$.

Así que $(f+g)^{p-1} \in \mathcal{L}^q$.

Por lo tanto:

$\left|\left|(f+g)^{p-1}\right|\right|_q = \left(\int_{\mathbb{F}} |f+g|^p \, d\mu\right)^{\frac{1}{q}} = \left(||f+g||_p\right)^{\frac{p}{q}}$,

$\int_{\mathbb{F}} |f+g|^p \, d\mu = \int_{\mathbb{F}} |f+g|^{p-1} \, |f+g| \, d\mu \leq \int_{\mathbb{F}} |f| \, |f+g|^{p-1} \, d\mu + \int_{\mathbb{F}} |g| \, |f+g|^{p-1} \, d\mu$

$\leq ||f||_p \left|\left|(f+g)^{p-1}\right|\right|_q + ||g||_p \left|\left|(f+g)^{p-1}\right|\right|_q$

$= \left(||f||_p + ||g||_p\right) \left(||f+g||_p\right)^{\frac{p}{q}}$.

Por lo tanto:

$||f+g||_p \leq \left(||f||_p + ||g||_p\right)^{\frac{1}{p}} \left(||f+g||_p\right)^{\frac{1}{q}}$.

Así que:

$\left(||f+g||_p\right)^{\frac{1}{p}} = \left(||f+g||_p\right)^{1-\frac{1}{q}} \leq \left(||f||_p + ||g||_p\right)^{\frac{1}{p}}$.

Si $p = \infty$, se tiene $|f+g| \leq |f| + |g| \leq \sup es\,(f) + \sup es\,(g)$ casi en todas partes, así que:

$||f+g||_\infty = \sup es\,(f+g) \leq \sup es\,(f) + \sup es\,(g) = ||f||_\infty + ||g||_\infty$.

■

Corolario 11.2. *Para cualquier $p \in [1, \infty]$, la función $||\,||_p$, definida sobre L^p, es una norma.*

Obsérvese que sobre $\mathcal{L}^p$, la función $||\,||_p$ no es una norma, pero es únicamente una propiedad de la norma la que no se cumple, a saber, que si $||f||_p = 0$ entonces $f = 0$. Si $||f||_p = 0$, únicamente se puede afirmar que $f = 0$ casi en todas partes.

Si $p \in (0, 1)$, la función $f \mapsto \left(\int_{\mathbb{F}} |f|^p \, d\mu\right)^{\frac{1}{p}}$ no es una norma. Ni siquiera lo es la función $x \mapsto \left(\sum_{k=1}^n |x_k|^p\right)^{\frac{1}{p}}$, definida sobre $\mathbb{R}^n$. Por ejemplo, en $\mathbb{R}^2$, si $x = (1,0)$ y $y = (0,1)$, se tiene:

$\left(\sum_{k=1}^{n}|x_k+y_k|^{\frac{1}{2}}\right)^2=4,$

$\left(\sum_{k=1}^{n}|x_k|^{\frac{1}{2}}\right)^2=1,$

$\left(\sum_{k=1}^{n}|y_k|^{\frac{1}{2}}\right)^2=1.$

Así que:

$\left(\sum_{k=1}^{n}|x_k+y_k|^{\frac{1}{2}}\right)^2>\left(\sum_{k=1}^{n}|x_k|^{\frac{1}{2}}\right)^2+\left(\sum_{k=1}^{n}|x_k+y_k|^{\frac{1}{2}}\right)^2.$

Por lo tanto, no se satisface la desigualdad del triángulo.

Si $p\in(0,1)$ y $f\in L^p$, definimos $||f||_p=\int_{\mathbb{F}}|f|^p\,d\mu$.

Lema 11.3. *Sean* $a,b\in[0,\infty)$ *y* $p\in(0,1)$, *entonces:*

$(a+b)^p\leq a^p+b^p$.

Demostración

Si $ab=0$, o $a=b$, la desigualdad es obvia.

Supongamos $a,b\in(0,\infty)$.

Definamos $f:[0,\infty)\mapsto\mathbb{R}$ de la siguiente manera:

$f(t)=1+t^p-(1+t)^p$.

Entonces, como $(1+t)^{1-p}>t^{1-p}$ para cualquier $t>0$,

$f'(t)=pt^{p-1}-p(1+t)^{p-1}>0$

para cualquier $t>0$.

Así que f es creciente en el intervalo $(0,\infty)$ y, siendo continua en el intervalo $[0,\infty)$, se tiene:

$f(t)\geq f(0)=0$.

Es decir:

$(1+t)^p\leq 1+t^p$

para cualquier $t>0$.

Por lo tanto:

$(a+b)^p=a^p\left(1+\frac{b}{a}\right)^p\leq a^p\left(1+\left(\frac{b}{a}\right)^p\right)=a^p+b^p.$

■

TEOREMA 11.3. *Para cualquier* $p \in (0,1)$*, la función* $||||_p$*, definida sobre* L^p*, es una norma.*

Demostración

$||f+g||_p = \int_{\mathbb{F}} |f+g|^p \, d\mu \leq \int_{\mathbb{F}} (|f|+|g|)^p \, d\mu \leq \int_{\mathbb{F}} |f|^p \, d\mu + \int_{\mathbb{F}} |g|^p \, d\mu$

$= ||f||_p + ||g||_p$.

■

PROPOSICIÓN 11.2. *Supongamos que la medida* μ *es finita, entonces, para cualquier* $r \in (0,\infty]$*, si* $f \in L^r$*, entonces* $f \in L^p$ *para cualquier* $p \in (0,r)$*.*

Demostración

Para $r \in (0,\infty)$, se tiene $y^p < 1 + y^r$ para cualquier $x \in [0,\infty)$ y cualquier $p \in (0,r)$, así que:

$\int_{\mathbb{F}} |f|^p \, d\mu \leq \mu(\mathbb{F}) + \int_{\mathbb{F}} |f|^r \, d\mu$.

Si $f \in \mathcal{L}^\infty$, se tiene $|f| \leq ||f||_\infty < \infty$ casi en todas partes. Así que

$\int_{\mathbb{F}} |f|^p \, d\mu \leq ||f||_\infty^p \, \mu(\mathbb{F}) < \infty$ para cualquier $p \in (0,\infty)$.

■

Si μ no es finita, el resultado anterior podría no ser válido. Por ejemplo, si $\mathbb{F} = [1,\infty)$, $\Im$ es la σ-álgebra formada por los conjuntos Lebesgue medibles y μ es la medida de Lebesgue, definamos $f : [1,\infty) \mapsto \mathbb{R}$ mediante la relación $f(x) = \frac{1}{x}$. Entonces $f \in L^2$, pero $f \notin L^1$.

PROPOSICIÓN 11.3. *Supongamos que existe una colección infinita numerable de conjuntos ajenos* A_n *de medida finita y positiva. Entonces, la dimensión de* L^p *es infinita.*

Demostración

Obviamente las funciones I_{A_n} pertenecen a L^p y son linealmente independientes.

■

11.2. Convergencia en L^p

DEFINICIÓN 11.1. *Para* $p \in (0,\infty]$*, se dice que una sucesión* $(f_n)_{n\in\mathbb{N}}$ *de funciones medibles converge en* L^p *a la función medible* f *si* $f, f_n \in L^p$ *para cualquier* $n \in \mathbb{N}$ *y*

$\text{lím}_{n \rightsquigarrow \infty} ||f_n - f||_p = 0$.

Si éste es el caso, se escribirá $f_n \xrightarrow{L^p} f$*.*

La convergencia en L^p tiene las propiedades comunes a la convergencia en cualquier espacio vectorial normado. En particular, si X es un espacio vectorial normado, sobre $\mathbb{R}$, con norma $\|\|$, se tiene lo siguiente:

(i) Si $(x_n)_{n\in\mathbb{N}}$ es una sucesión que converge x, entonces cualquier subsucesión de $(x_n)_{n\in\mathbb{N}}$ también converge a x.
(ii) Si $(x_n)_{n\in\mathbb{N}}$ es una sucesión de elementos de X tales que $x_n \longrightarrow x$ y $x_n \longrightarrow y$, entonces $x = y$.
(iii) Si $c \in \mathbb{R}$ y $(x_n)_{n\in\mathbb{N}}$ es una sucesión de elementos de X tales que $x_n \longrightarrow x$, entonces $cx_n \longrightarrow cx$.
(iv) Si $(x_n)_{n\in\mathbb{N}}$ y $(y_n)_{n\in\mathbb{N}}$ son dos sucesiones de elementos de X tales que $x_n \longrightarrow x$ y $y_n \longrightarrow y$, entonces $x_n + y_n \longrightarrow x + y$.
(v) Si $(x_n)_{n\in\mathbb{N}}$ es una sucesión de elementos de X tales que la serie $\sum_{n=1}^{\infty} \|x_n\|$ converge, entonces la sucesión $(y_n)_{n\in\mathbb{N}}$ definida por $y_n = \sum_{k=1}^{n} x_k$ es de Cauchy.
(vi) Si para cualquier sucesión $(x_n)_{n\in\mathbb{N}}$ de elementos de X tales que la serie $\sum_{n=1}^{\infty} \|x_n\|$ converge, la sucesión $(y_n)_{n\in\mathbb{N}}$, definida por $y_n = \sum_{k=1}^{n} y_k$, converge. Entonces X es completo.

TEOREMA 11.4. *Para cualquier $p \in (0, \infty]$, L^p, con la norma $\|\|_p$ es un espacio normado completo.*

Demostración

Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones en L^p tales que la serie $\sum_{n=1}^{\infty} ||f_n||_p$ converge y, para cada $n \in \mathbb{N}$, definamos $h_n = \sum_{k=1}^{n} f_k$.

Definamos:

$$f(x) = \begin{cases} \sum_{k=1}^{\infty} f_k(x) & \text{si } \sum_{k=1}^{\infty} |f_k(x)| < \infty \\ 0 & \text{en otro caso} \end{cases}$$

y, para cada $n \in \mathbb{N}$, $g_n = \sum_{k=1}^{n} |f_k|$.

Para cada $x \in \mathbb{F}$, $(g_n(x))_{n\in\mathbb{N}}$ es una sucesión no decreciente. Sea $g(x) = \lim_{n\rightsquigarrow\infty} g_n(x)$.

Por el lema de Fatou, se tiene:

$$\left(\int_{\mathbb{F}} g^p d\mu\right)^{\frac{1}{p}} \leq \lim_{n\rightsquigarrow\infty} \left(\int_{\mathbb{F}} g_n^p d\mu\right)^{\frac{1}{p}}$$

$$= \lim_{n\rightsquigarrow\infty} ||\sum_{k=1}^{n} |f_k|||_p \leq \lim_{n\rightsquigarrow\infty} \sum_{k=1}^{n} ||f_k||_p < \infty$$

para $p \in [1, \infty)$.

$$\int_{\mathbb{F}} g^p d\mu \leq \lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} g_n^p d\mu = \lim_{n\rightsquigarrow\infty} ||\sum_{k=1}^{n} |f_k|||_p$$

$$\leq \lim_{n\rightsquigarrow\infty} \sum_{k=1}^{n} ||f_k||_p < \infty$$

para $p \in (0, 1)$.

Así que, para cualquier $p \in (0, \infty)$, $g \in L^p$.

Además:

$\left|\sum_{k=1}^{n} f_k\right| \leq \sum_{k=1}^{n} |f_k| \leq g$.

Así que $|f| \leq g$. En particular $f \in L^p$.

Por otra parte:

$\left|\sum_{k=1}^{n} f_k - f\right|^p \leq \left(\left|\sum_{k=1}^{n} f_k\right| + |f|\right)^p \leq (2g)^p$.

Así que, por el teorema de la convergencia dominada:

$\text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} \left|\sum_{k=1}^{n} f_k - f\right|^p = \int_{\mathbb{F}} \text{lím}_{n\rightsquigarrow\infty} \left|\sum_{k=1}^{n} f_k - f\right|^p = 0$.

Si $p = \infty$, se tiene:

$|f| = \text{lím}_{n\rightsquigarrow\infty} \left|\sum_{k=1}^{n} f_k\right| \leq \text{lím}_{n\rightsquigarrow\infty} \sum_{k=1}^{n} |f_k| \leq \text{lím}_{n\rightsquigarrow\infty} \sum_{k=1}^{n} ||f_k||_\infty$

$= \sum_{k=1}^{\infty} ||f_k||_\infty < \infty$

casi en todas partes.

Así que $f \in L^\infty$.

Dada $\varepsilon > 0$, tomemos $N \in \mathbb{N}$ tal que $\sum_{k=N}^{\infty} ||f_k||_\infty < \varepsilon$. Entonces, para cualquier $n \geq N$, se tiene:

$\left|f - \sum_{k=1}^{n} f_k\right| = \text{lím}_{m\rightsquigarrow\infty} \left|\sum_{k=n+1}^{m} f_k(x)\right| \leq \text{lím}_{m\rightsquigarrow\infty} \sum_{k=n+1}^{m} |f_k(x)|$

$\leq \sum_{k=n+1}^{\infty} ||f_k||_p$

casi en todas partes.

Así que:

$\left|\left|\sum_{k=1}^{n} f_k - f\right|\right|_\infty \leq \sum_{k=n+1}^{\infty} ||f_k||_\infty < \varepsilon$.

Por lo tanto, en cualquier caso, la sucesión $(h_n)_{n\in\mathbb{N}}$ converge en L^p. Así que L^p es completo. ∎

Teorema 11.5. *Sea $p \in (0, \infty]$ y $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{L^p} f$, entonces $f_n \xrightarrow{\mu} f$.*

Demostración

Sea $p \in (0, \infty)$.

Para cualquier $\varepsilon > 0$, se tiene:

$\int_{\mathbb{F}} |f_n - f|^p \, d\mu = \int_{\{|f_n - f| > \varepsilon\}} |f_n - f|^p \, d\mu + \int_{\{|f_n - f| \leq \varepsilon\}} |f_n - f|^p \, d\mu$

$\geq \int_{\{|f_n - f| > \varepsilon\}} |f_n - f|^p \, d\mu \geq \varepsilon^p \mu \left[|f_n. - f| > \varepsilon \right]$.

Así que:

$\mu \left[|f_n. - f| > \varepsilon \right] \leq \frac{1}{\varepsilon^p} \int_{\mathbb{F}} |f_n - f|^p \, d\mu$.

Por lo tanto, $\lim_{n \rightsquigarrow \infty} \mu \left[|f_n. - f| > \varepsilon \right] = 0$ para cualquier $\varepsilon > 0$.

Supongamos que $f_n \xrightarrow{L^\infty} f$, entonces $\lim_{n \rightsquigarrow \infty} \|f_n. - f\|_\infty = 0$ y, para cualquier $n \in \mathbb{N}$, $|f_n. - f| \leq \|f_n. - f\|_\infty$ casi en todas partes.

Dada $\varepsilon > 0$, sea N tal que $\|f_n. - f\|_\infty < \varepsilon$ para cualquier $n \geq N$. Entonces , para cualquier $n \geq N$, $|f_n. - f| \leq \|f_n. - f\|_\infty < \varepsilon$ casi en todas partes. Así que $\mu \left[|f_n. - f| > \varepsilon \right] = 0$ para cualquier $n \geq N$. Por lo tanto, $\lim_{n \rightsquigarrow \infty} \mu \left[|f_n. - f| > \varepsilon \right] = 0$ para cualquier $\varepsilon > 0$.

■

Como se muestra en el siguiente ejemplo, el inverso del resultado anterior no es válido en general.

Ejemplo 11.1. *Sea $\mathbb{F} = (0, 1]$, μ la medida de Lebesgue sobre $\mathbb{F}$, $p \in (0, \infty)$ y, para $n \in \{0, 1, 2, \ldots\}$ y $j \in \{0, 1, 2, \ldots, 2^n - 1\}$, $f_{2^n+j} = (2^n)^{\frac{1}{p}} I_{(\frac{j}{2^n}, \frac{j+1}{2^n}]}$.*

Para cualquier $n \in \{0, 1, 2, \ldots\}$ y $j \in \{0, 1, 2, \ldots, 2^n - 1\}$, f_{2^n+j} toma únicamente los valores 0 y $(2^n)^{\frac{1}{p}}$, y $\mu\left(\left\{y \in \mathbb{F} : f_{2^n+j}(y) = (2^n)^{\frac{1}{p}}\right\}\right) = \frac{1}{2^n}$. Así que $\lim_{n \rightsquigarrow \infty} \mu \left[|f_n| > \varepsilon \right] = 0$ para cualquier $\varepsilon > 0$. Por lo tanto, $f_n \xrightarrow{\mu} 0$.

Por otra parte, para cualquier $n \in \{0, 1, 2, \ldots\}$, se tiene:

$\int_{\mathbb{F}} |f_{2^{n+1}} - f_{2^n}|^p \, d\mu = \int_{\mathbb{F}} \left| (2^{n+1})^{\frac{1}{p}} I_{(0, \frac{1}{2^{n+1}}]} - (2^n)^{\frac{1}{p}} I_{(0, \frac{1}{2^n}]} \right|^p d\mu$

$\geq \int_{(\frac{1}{2^{n+1}}, \frac{1}{2^n}]} \left| (2^{n+1})^{\frac{1}{p}} I_{(0, \frac{1}{2^{n+1}}]} - (2^n)^{\frac{1}{p}} I_{(0, \frac{1}{2^n}]} \right|^p d\mu$

$= \int_{(\frac{1}{2^{n+1}}, \frac{1}{2^n}]} \left| (2^n)^{\frac{1}{p}} I_{(0, \frac{1}{2^n}]} \right|^p d\mu = \int_{\mathbb{F}} \left| (2^n)^{\frac{1}{p}} I_{(\frac{1}{2^{n+1}}, \frac{1}{2^n}]} \right|^p d\mu$

$= 2^n \left(\frac{1}{2^n} - \frac{1}{2^{n+1}} \right) = \frac{1}{2}$.

Así que la sucesión $(f_n)_{n \in \mathbb{N}}$ no es de Cauchy en L^p y, por lo tanto, no converge en L^p.

Para $p = \infty$ tomemos $f_{2^n+j} = I_{(\frac{j}{2^n}, \frac{j+1}{2^n}]}$, entonces, como se mostró en el ejemplo 10.2, se tiene $f_n \xrightarrow{\mu} 0$. Por otra parte, para cualquier $n \in \{0, 1, 2, \ldots\}$, se tiene:

$||f_{2^{n+1}} - f_{2^n}||_\infty = 1$.

Así que la sucesión $(f_n)_{n\in\mathbb{N}}$ *no es de Cauchy en* L^∞ *y, por lo tanto, no converge en* L^∞*.*

TEOREMA 11.6. *Sean* $p \in (0,\infty)$, $\{f_n\}_{n\in\mathbb{N}}$ *una sucesión en* L^p *y* f *una función medible. Si* μ *es una medida finita, las siguientes propiedades son equivalentes:*

(i) $f_n \xrightarrow{\mu} f$ *y la familia* $\{|f_n|^p : n \in \mathbb{N}\}$ *es uniformemente integrable.*
(ii) *ii)* $f \in L^p$ *y* $f_n \xrightarrow{L^p} f$.
(iii) *iii)* $f \in L^p$, $f_n \xrightarrow{\mu} f$ *y* $\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n|^p \, d\mu = \int_{\mathbb{F}} |f|^p \, d\mu$.

Demostración

$i \Rightarrow ii$

Por la proposición 10.6 existe una subsucesión $(f_{n_k})_{k\in\mathbb{N}}$ tal que $f_{n_k} \xrightarrow{c.t.p.} f$. Así que, por la proposición 8.5, $|f|^p$ es integrable.

Por otra parte:

$|f_n - f|^p \leq 2^p \left(|f_n|^p + |f|^p\right)$.

Así que la familia $\{|f_n - f|^p\}_{n\in\mathbb{N}}$ también es uniformemente integrable.

Para cada $\alpha > 0$, se tiene:

$\int_{\mathbb{F}} |f_n - f|^p \, d\mu = \int_{\{|f_n - f|^p > \alpha\}} |f_n - f|^p \, d\mu + \int_{\{|f_n - f|^p \leq \alpha\}} |f_n - f|^p \, d\mu$

$\leq \int_{\{|f_n - f|^p > \alpha\}} |f_n - f|^p \, d\mu + \alpha\mu(\mathbb{F})$.

Pero, $\lim_{n\rightsquigarrow\infty} \mu\left[|f_n - f|^p > \alpha\right] = \lim_{n\rightsquigarrow\infty} \mu\left[|f_n - f| > \alpha^{\frac{1}{p}}\right] = 0$ y la familia $\{|f_n - f|^p\}_{n\in\mathbb{N}}$ es uniformemente integrable, así que:

$\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|^p \, d\mu \leq \lim_{n\rightsquigarrow\infty} \int_{\{|f_n - f|^p > \alpha\}} |f_n - f|^p \, d\mu + \alpha\mu(\mathbb{F}) = \alpha\mu(\mathbb{F})$.

Así que, como $\alpha > 0$ es arbitraria, se tiene:

$\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|^p \, d\mu = 0$.

$ii \Rightarrow iii$

Para todo espacio vectorial normado $(\mathbb{X}, ||||)$, se tiene:

$|||x|| - ||y||| \leq ||x - y||$.

Así que, si una sucesión $(x_n)_{n\in\mathbb{N}}$ converge a x, entonces:

$|||x_n|| - ||x||| \leq ||x_n - x||$.

Por lo tanto:

$\text{lím}_{n\rightsquigarrow\infty}\left(\left|||x_n||-||x||\right|\right)\leq \text{lím}_{n\rightsquigarrow\infty}||x_n-x||=0.$

Se concluye entonces que:

$\text{lím}_{n\rightsquigarrow\infty}||x_n||=||x||.$

Si $f\in L^p$ y $f_n\xrightarrow{L^p}f$, entonces:

$\text{lím}_{n\rightsquigarrow\infty}\left(\int_{\mathbb{F}}|f_n|^p\,d\mu\right)^{\frac{1}{p}}=\left(\int_{\mathbb{F}}|f|^p\,d\mu\right)^{\frac{1}{p}}$ si $p\in[1,\infty)$,

$\text{lím}_{n\rightsquigarrow\infty}\int_{\mathbb{F}}|f_n|^p\,d\mu=\int_{\mathbb{F}}|f|^p\,d\mu$ si $p\in(0,1)$.

Así que, en cualquier caso:

$\text{lím}_{n\rightsquigarrow\infty}\int_{\mathbb{F}}|f_n|^p\,d\mu=\int_{\mathbb{F}}|f|^p\,d\mu.$

$iii\Rightarrow i$

Para $\beta>0$ y $x\in\mathbb{R}$, definamos:

$$f_\beta(x)=\begin{cases}|x|^p & \text{si } |x|\leq\beta^{\frac{1}{p}}\\ -\frac{\beta}{(2\beta)^{\frac{1}{p}}-\beta^{\frac{1}{p}}}\left(x-(2\beta)^{\frac{1}{p}}\right) & \text{si } x>0 \text{ y } \beta^{\frac{1}{p}}<|x|<(2\beta)^{\frac{1}{p}}\\ \frac{\beta}{(2\beta)^{\frac{1}{p}}-\beta^{\frac{1}{p}}}\left(x+(2\beta)^{\frac{1}{p}}\right) & \text{si } x<0 \text{ y } \beta^{\frac{1}{p}}<|x|<(2\beta)^{\frac{1}{p}}\\ 0 & \text{si } |x|\geq(2\beta)^{\frac{1}{p}}\end{cases}$$

f_β es una función continua, así que $f_\beta\circ f_n\xrightarrow{\mu}f_\beta\circ f$.

Además, $f_\beta(x)\leq|x|^p$ y $0\leq f_\beta(x)\leq\beta$ para cualquier $x\in\mathbb{R}$.

Se tiene $|f_\beta\circ f_n|\leq\beta$ y $f_\beta\circ f_n\xrightarrow{\mu}f_\beta\circ f$, así que, por la proposición 10.9:

$\text{lím}_{n\rightsquigarrow\infty}\int_{\mathbb{F}}|f_\beta\circ f_n-f_\beta\circ f|\,d\mu=0.$

Por lo tanto:

$\text{lím}_{n\rightsquigarrow\infty}\int_{\mathbb{F}}f_\beta\circ f_n d\mu=\int_{\mathbb{F}}f_\beta\circ f d\mu.$

Se tiene:

$\text{lím inf}_{n\rightsquigarrow\infty}\int_{\{y\in\mathbb{F}:|f_n(y)|^p\leq 2\beta\}}|f_n|^p\,d\mu\geq\int_{\{y\in\mathbb{F}:|f_n(y)|^p\leq 2\beta\}}f_\beta\circ f_n d\mu=\int_{\mathbb{F}}f_\beta\circ f_n d\mu.$

Así que:

$\text{lím inf}_{n\rightsquigarrow\infty}\int_{\{y\in\mathbb{F}:|f_n(y)|^p\leq 2\beta\}}|f_n|^p\,d\mu\geq\text{lím}_{n\rightsquigarrow\infty}\int_{\mathbb{F}}f_\beta\circ f_n d\mu=\int_{\mathbb{F}}f_\beta\circ f d\mu$

$\geq\int_{\{y\in\mathbb{F}:|f(y)|^p\leq\beta\}}f_\beta\circ f d\mu=\int_{\{y\in\mathbb{F}:|f(y)|^p\leq\beta\}}|f|^p\,d\mu.$

Además, por hipótesis:

$\lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n|^p \, d\mu = \int_{\mathbb{F}} |f|^p \, d\mu$.

Por lo tanto:

$\limsup_{n\rightsquigarrow\infty} \int_{\{y\in\mathbb{F}:|f_n(y)|^p>2\beta\}} |f_n|^p \, d\mu = \limsup_{n\rightsquigarrow\infty} \left(\int_{\mathbb{F}} |f_n|^p \, d\mu - \int_{\{y\in\mathbb{F}:|f_n(y)|^p\leq 2\beta\}} |f_n|^p \, d\mu \right)$

$= \lim_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n|^p \, d\mu - \liminf_{n\rightsquigarrow\infty} \int_{\{y\in\mathbb{F}:|f_n(y)|^p\leq 2\beta\}} |f_n|^p \, d\mu$

$\leq \int_{\mathbb{F}} |f|^p \, d\mu - \int_{\{y\in\mathbb{F}:|f(y)|^p\leq \beta\}} |f|^p \, d\mu = \int_{\{y\in\mathbb{F}:|f(y)|^p> \beta\}} |f|^p \, d\mu$.

Dada $\varepsilon > 0$, tomemos β_0 tal que $\int_{\{y\in\mathbb{F}:|f(y)|^p>\beta_0\}} |f|^p \, d\mu < \varepsilon$. Entonces:

$\lim_{n\rightsquigarrow\infty} \left(\sup \left\{ \int_{\{y\in\mathbb{F}:|f_j(y)|^p>2\beta_0\}} |f_j|^p \, d\mu : j \geq n \right\} \right)$

$= \limsup_{n\rightsquigarrow\infty} \int_{\{y\in\mathbb{F}:|f_n(y)|^p>2\beta_0\}} |f_n|^p \, d\mu \leq \int_{\{y\in\mathbb{F}:|f(y)|^p>\beta_0\}} |f|^p \, d\mu < \varepsilon$.

Así que existe $N \in \mathbb{N}$ tal que:

$\sup \left\{ \int_{\{y\in\mathbb{F}:|f_j(y)|^p>2\beta_0\}} |f_j|^p \, d\mu : j \geq n \right\} < \varepsilon$

para cualquier $n \geq N$.

Para cada $j \in \{1, 2, \ldots, N-1\}$ tomemos β_j tal que:

$\int_{\left\{y\in\mathbb{F}:|f_j(y)|^p>2\beta_j\right\}} |f_j|^p \, d\mu < \varepsilon$

Definiendo $\alpha = \max\left\{2\beta_0, 2\beta_1, \ldots, 2\beta_{N-1}\right\}$, se tiene:

$\sup \left\{ \int_{\{y\in\mathbb{F}:|f_j(y)|^p>\alpha\}} |f_j|^p \, d\mu : j \in \mathbb{N} \right\} < \varepsilon$

Por lo tanto:

$\lim_{\alpha\rightsquigarrow\infty} \sup \left\{ \int_{\{y\in\mathbb{F}:|f_j(y)|^p>\alpha\}} |f_j|^p \, d\mu : j \in \mathbb{N} \right\} = 0$.

Así que la familia $\{|f_n|^p : n \in \mathbb{N}\}$ es uniformemente integrable.

∎

Proposición 11.4. *Supongamos que la medida μ es finita, entonces, para cualquier $r \in (0, \infty]$, si $f_n \xrightarrow{L^r} f$, entonces $f_n \xrightarrow{L^p} f$ para cualquier $p \in (0, r]$.*

Demostración

Si $f_n \xrightarrow{L^\infty} f$, entonces $\text{lím}_{n\rightsquigarrow\infty} \sup es \{|f_n - f|\} = 0$. Así que, dado $\varepsilon > 0$, existe N tal que $\sup es \{|f_n - f|\} < \varepsilon$ para cualquier $n \geq N$. Por consiguiente, $|f_n - f| < \varepsilon$ casi en todas partes. Así que $\int_{\mathbb{F}} |f_n - f|^p \, d\mu \leq \varepsilon^p \mu(\mathbb{F})$ para cualquier $n \geq N$ y $p \in (0, \infty)$. Por lo tanto:

$\text{lím}_{n\rightsquigarrow\infty} \int_{\mathbb{F}} |f_n - f|^p \, d\mu = 0$.

Si $r \in (0, \infty)$ y $f_n \xrightarrow{L^r} f$, entonces $f_n \xrightarrow{\mu} f$ y la familia $\{|f_n|^r\}_{n\in\mathbb{N}}$ es uniformemente integrable.

Además, $|y|^p \leq 1 + |y|^r$ para cualquier $y \in \mathbb{R}$ y cualquier $p \in (0, r]$. Por lo tanto, para $\alpha > 2$, se tiene:

$$\int_{\{|f_n|^p > \alpha\}} |f_n|^p \, d\mu \leq \int_{\{1+|f_n|^r > \alpha\}} (1 + |f_n|^r) \, d\mu \leq 2 \int_{\{|f_n|^r > \alpha - 1\}} |f_n|^r \, d\mu.$$

Así que:

$$\text{lím}_{\alpha\rightsquigarrow\infty} \sup \left\{ \int_{\{|f_n|^p > \alpha\}} |f_n|^p \, d\mu : n \in \mathbb{N} \right\}$$

$$\leq 2 \, \text{lím}_{\alpha\rightsquigarrow\infty} \sup \left\{ \int_{\{|f_n|^r > \alpha - 1\}} |f_n|^r \, d\mu : n \in \mathbb{N} \right\} = 0.$$

Por lo tanto, la familia $\{|f_n|^p\}_{n\in\mathbb{N}}$ es uniformemente integrable.

Finalmente, como también se tiene $f_n \xrightarrow{\mu} f$, entonces $f \in L^p$ y $f_n \xrightarrow{L^p} f$.

■

Como se muestra en el siguiente ejemplo, en general, para $p \in (0, \infty)$, la convergencia en L^p no implica convergencia casi en todas partes.

Ejemplo 11.2. *Consideremos la sucesión de funciones $(f_n)_{n\in\mathbb{N}}$ del ejemplo 10.2. Para cualquier $p \in (0, \infty)$, $n \in \{0, 1, 2, \ldots\}$ y $j \in \{0, 1, 2, \ldots, 2^n - 1\}$, se tiene:*

$$\int_{\mathbb{F}} |f_{2^n+j}|^p \, d\mu = \int_{\mathbb{F}} \left| I_{\left(\frac{j}{2^n}, \frac{j+1}{2^n}\right]} \right|^p d\mu = \frac{1}{2^n}.$$

Así que, $f_n \xrightarrow{L^p} 0$, pero, como vimos, para cualquier $x \in \mathbb{F}$, la sucesión $(f_n(x))_{n\in\mathbb{N}}$ no converge.

Proposición 11.5. *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{L^\infty} f$, entonces $f_n \xrightarrow{c.t.p} f$.*

Demostración

Para cada $n \in \mathbb{N}$, definamos:

$c_n = ||f_n - f||_\infty$.

Por la definición de la norma $||||_\infty$, para cada $n \in \mathbb{N}$ existe un conjunto $A_n \in \Im$ tal que $\mu(A_n) = 0$ y:

$|f_n(x) - f(x)| \leq c_n$

para cualquier $x \notin A_n$.

Definamos $A = \bigcup_{n=1}^{\infty} A_n$, entonces $\mu(A) = 0$ y:

$|f_n(x) - f(x)| \leq c_n$

para cualquier $x \notin A$ y cualquier $n \in \mathbb{N}$.

Como $f_n \xrightarrow{L^\infty} f$, la sucesión $(c_n)_{n\in\mathbb{N}}$ converge a cero. Así que, para cualquier $x \notin A$, se tiene:

$\text{lím sup}_{n\rightsquigarrow\infty} |f_n(x) - f(x)| \leq \text{lím}_{n\rightsquigarrow\infty} c_n = 0.$

Por lo tanto, $\text{lím}_{n\rightsquigarrow\infty} |f_n(x) - f(x)|$. Así que $f_n \xrightarrow{c.t.p} f$.

■

Como se muestra en el siguiente ejemplo, en general, para $p \in (0, \infty]$, la convergencia casi en todas partes no implica la convergencia en L^p.

Ejemplo 11.3. *Sea $\mathbb{F} = (0, 1]$, μ la medida de Lebesgue sobre $\mathbb{F}$ y $p \in (0, \infty)$. Para cada $n \in \mathbb{N}$, definamos $f_n : \mathbb{F} \to \mathbb{R}$ de la siguiente manera:*

$$f_n(x) = \begin{cases} (2^n)^{\frac{1}{p}} & si\ x \in \left(0, \frac{1}{2^n}\right) \\ 0 & en\ otro\ caso \end{cases}$$

Para cualquier $x \in \mathbb{F}$ se tiene $\text{lím}_{n\to\infty} f_n(x) = 0$.

Por otra parte, para cualquier $n \in \mathbb{N}$, se tiene:

$\int_{\mathbb{F}} |f_{n+1} - f_n|^p \, d\mu \geq \int_{\left(\frac{1}{2^{n+1}}, \frac{1}{2^n}\right]} |f_{n+1} - f_n|^p \, d\mu = \int_{\left(\frac{1}{2^{n+1}}, \frac{1}{2^n}\right]} |f_n|^p \, d\mu = 2^n \left(\frac{1}{2^n} - \frac{1}{2^{n+1}}\right) = \frac{1}{2}.$

Así que la sucesión $(f_n)_{n\in\mathbb{N}}$ no es de Cauchy en L^p y, por lo tanto, no converge en L^p.

Si $p = \infty$ tomemos:

$$f_n(x) = \begin{cases} 1 & si\ x \in \left(0, \frac{1}{2^n}\right) \\ 0 & en\ otro\ caso \end{cases}$$

Para cualquier $x \in \mathbb{F}$ se tiene $\text{lím}_{n\to\infty} f_n(x) = 0$.

Por otra parte, para cualquier $n \in \mathbb{N}$, se tiene:

$||f_{n+1} - f_n||_\infty = 1.$

Así que la sucesión $(f_n)_{n\in\mathbb{N}}$ no es de Cauchy en L^∞ y, por lo tanto, no converge en L^∞.

11.3. Densidad de las funciones simples en L^p

Lema 11.4. *Sea $p \in (0, \infty)$. Una función simple no negativa φ pertenece a L^p si sólo si φ es nula fuera de un conjunto de medida finita.*

Demostración

Sea φ una función simple no negativa con representación canónica $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$. Entonces, para $p \in (0, \infty)$, se tiene:

$\int_{\mathbb{F}} \varphi^p d\mu = \sum_{k=1}^{n} a_k^p \mu(E_k)$.

Así que $\varphi \in L^p$ si y sólo si $\mu\left(\bigcup_{k=1}^{n} E_k\right) = \sum_{k=1}^{n} \mu(E_k) < \infty$.

Además:

$\{x \in \mathbb{F} : \varphi(x) > 0\} = \bigcup_{k=1}^{n} I_{E_k}$.

Por lo tanto, $\varphi \in L^p$ si y sólo si $\mu\left(\{x \in \mathbb{F} : \varphi(x) > 0\}\right) < \infty$.

$\blacksquare$

Teorema 11.7. *Sea $p \in (0, \infty)$. El conjunto de las funciones simples nulas fuera de un conjunto de medida finita es denso en L^p.*

Demostración

Sea $f \in L^p$ y $(\varphi_n)_{n\in\mathbb{N}}$ y $(\psi_n)_{n\in\mathbb{N}}$ dos sucesiones no decrecientes de funciones simples no negativas tales que $\lim_{n\to\infty} \varphi_n(x) = f^+(x)$ y $\lim_{n\rightsquigarrow\infty} \psi_n(x) = f^-(x)$ para cualquier $x \in \mathbb{F}$. Sea $A = \{y \in : |f(y)| < \infty\}$. Entonces, para cualquier $n \in \mathbb{N}$, $\varphi_n I_A$ y $\psi_n I_A$ siguen siendo simples.

Para cualquier $n \in \mathbb{N}$ se tiene $\varphi_n I_A \leq f^+$ y $\psi_n I_A \leq f^-$, así que $\varphi_n I_A \in L^p$ y $\psi_n I_A \in L^p$. Por lo tanto, $\varphi_n I_A$ y $\psi_n I_A$ son nulas fuera de un conjunto de medida finita.

Además, como $\mu(A^c) = 0$, $\lim_{n\to\infty} \varphi_n I_A = f^+$ y $\lim_{n\rightsquigarrow\infty} \psi_n I_A = f^-$ casi en todas partes.

Para cada $n \in \mathbb{N}$ definamos $s_n = \varphi_n I_A - \psi_n I_A$. Entonces:

$\lim_{n\rightsquigarrow\infty} s_n = f^+ - f^- = f$ casi en todas partes. Así que:

$\lim_{n\rightsquigarrow\infty} |f - s_n|^p = 0$ casi en todas partes.

Además:

$|s_n| = |\varphi_n I_A - \psi_n I_A| \leq \varphi_n + \psi_n = |f|$

Así que:

$|f - s_n|^p \leq 2^p \left(|f|^p + |s_n|^p\right) \leq 2^{p+1} |f|^p$.

Por lo tanto, aplicando el teorema de la convergencia dominada, se tiene:

$\lim_{n \rightsquigarrow \infty} \int_{\mathbb{F}} |f - s_n|^p \, d\mu = 0.$

Así que la sucesión $(s_n)_{n \in \mathbb{N}}$ converge a f en L^p.

■

CAPÍTULO 12

TEORÍA DE LA PROBABILIDAD
Desarrollo histórico

El Cálculo de Probabilidades tuvo su origen en el estudio de algunos problemas relacionados con juegos de dados. La dificultad y al mismo tiempo lo interesante de estos juegos consiste en que al lanzar uno o varios dados, de la manera usual en que esto se realiza, el resultado que se obtiene no está únicamente determinado, sino que es uno de un conjunto de posibles resultados. La teoría se fue desarrollando con la solución de problemas cada vez más complejos y de diversa índole, no únicamente en el ámbito de los juegos con dados, sino abarcando incluso fenómenos naturales.

En situaciones como la del lanzamiento de varios dados, donde existen diferentes posibles resultados, el resultado es únicamente uno de ellos, pero, a priori, no se sabe cuál; de ahí que podemos hacer aseveraciones acerca del resultado que obtendremos, las cuales tienen la característica de que pueden resultar ser verdaderas o falsas una vez que observamos el resultado obtenido. Por ejemplo, al lanzar tres dados, la aseveración " la suma de los números que muestran las caras superiores de los dados, una vez lanzados, es mayor que 10" puede resultar verdadera o falsa cuando se realiza el lanzamiento. A cada una de esas aseveraciones que se pueden hacer acerca del resultado que se obtiene se le llama un **evento**; cuando la aseveración resulta ser verdadera, decimos que el evento ocurre; cuando resulta falsa, decimos que no ocurre.

Considerando que un evento puede ocurrir o no ocurrir, lo que se busca es asignar un número real no negativo a cada evento, el cual indique que tanto se puede esperar que el evento ocurra. A ese número se le llama la probabilidad del evento. De esta forma, la probabilidad la podemos ver como una función no negativa definida sobre la familia de eventos, a la cual llamaremos **función de probabilidad**.

Para fines de aplicar el Cálculo de Probabilidades se introduce el concepto de **experimento aleatorio**, donde un experimento se considera como cualquier proceso que conduzca a un resultado. Algunos experimentos pueden consistir simplemente en la observación de un determinado fenómeno natural y en ir anotando algunas de las características que observamos;

por ejemplo, la medición día con día del crecimiento de una planta. En otros experimentos están involucradas acciones nuestras, además de la observación, como en el caso del lanzamiento de un dado.

Un experimento consta de dos partes, por un lado tenemos la descripción del proceso que estamos considerando, especificando las condiciones bajo las cuales se realiza, y por otro lado tenemos lo que consideramos como su resultado.

> Se dice que un experimento es aleatorio cuando al realizarse, bajo las condiciones que se indiquen, su resultado puede ser cualquiera de un conjunto de resultados posibles. Cabe aclarar que la posibilidad de distintos resultados puede provenir de que las condiciones que se especifican para la realización del experimento incluyen cierta arbitrariedad.

En la formulación moderna, para modelar matematicamente un experimento aleatorio se busca construir un espacio de medida $(\Omega, \Im, P)$, donde Ω, llamado el espacio muestral del experimento, es el conjunto de sus posibles resultados, $\Im$ es una σ-álgebra de subconjuntos de Ω, los cuales representan a los eventos, y la medida P, definida sobre $\Im$, tiene la característica de asignarle a Ω el valor 1. Cada elemento de $\Im$ es entonces un subconjunto del total de posibles resultados del experimento aleatorio en consideración y lo que nos interesa obtener es la probabilidad de que ocurra alguno de los posibles resultados que pertenecen a ese subconjunto. En general, la medida P se construye mediante un proceso de extensión: se comienza por asignar probabilidades a algunos eventos y después, utilizando las propiedades de P, se busca extenderla hasta abarcar el mayor número posible de subconjuntos de Ω. De esta forma, el modelo matemático para el estudio del experimento consta de 3 elementos, el primero es el conjunto Ω, el segundo una familia de subconjuntos de Ω y el tercero una medida, que llamaremos medida de probabilidad, la cual asocia a cada elemento A de $\Im$ un número real que designa la probabilidad de que ocurra alguno de los posibles resultados que pertenecen a A.

Este modelo se fue construyendo durante los primeros 33 años del siglo XX, buscando axiomatizar el Cálculo de Probabilidades. Durante esos 33 años se estableció y consolidó su vínculo con la Teoría de la Medida, la cual surgió en los primeros años del siglo con los trabajos de Émile Borel y Henri Lebesgue. Es con su formulación axiomática que podemos hablar de una teoría matemática de la probabilidad, ya que de esta forma se tiene un cuerpo teórico, independiente de los problemas reales específicos que se tratan con el Cálculo de Probabilidades.

La identificación de una función de probabilidad con una medida no surgió de manera automática como algo general aplicable a cualquier fenómeno aleatorio, sino que requirió de un proceso, que llevó varios años, en el cual se mostró que es una alternativa adecuada. En el centro de este proceso se encuentra el planteamiento de problemas donde se trata de calcular probabilidades de eventos cuya ocurrencia o no ocurrencia depende de una infinidad de observaciones y la aceptación de la σ-aditividad como una propiedad de cualquier función de probabilidad, la cual permite atacar ese tipo de problemas.

12.1. Origen del Cálculo de Probabilidades

Los juegos donde se utilizan dados o algo similar se inventaron hace miles de años; el dado más antiguo que se conoce fue encontrado en Irak y tiene una antigüedad de alrededor de 5 mil años. El astragalo, también conocido como tali o taba, es un precursor del dado. Se trata de un hueso del pie de un animal, el cual presenta cuatro caras, cada una con un nombre que más tarde se convirtió en un número. Las más utilizadas son las del carnero. Los juegos con tabas son tan antiguos como los juegos con dados; se han encontrado tabas en las tumbas de los faraones egipcios. Incluso hoy en día hay lugares, incluyendo algunas regiones de México, donde se juega con una taba.

En la antigua Grecia y en Roma se volvió muy popular un juego que consistía en lanzar 4 tabas; en las caras de cada una de ellas estaban marcados los números 1, 3, 4 y 6, respectivamente. Las diferentes combinaciones tenían diferentes valores, siendo considerada la más alta la que consiste en 4 números distintos, la cual era llamada Venus.

Los aztecas y otros pueblos originarios de América utilizaban algo que puede considerarse equivalente a los dados; se trata de unos granos rojos parecidos al frijol, cada uno de los cuales se pintaba de un lado. Esos granos son semillas de un árbol que era conocido como tzité y a los granos mismos se les llamaba tzité. Actualmente se les conoce como colorines. Los granos eran utilizados para fines adivinatorios y para jugar un juego que llamaban patolli, el cual se jugaba sobre un tablero en forma de cruz diagonal, dividida en 52 casillas y se utilizaban 5 colorines marcados cada uno de un lado. Cada jugador colocaba una apuesta (había quienes apostaban incluso su persona y si perdían quedaban sometidos a la condición de esclavos). Ganaba el juego quien lograra primero recorrer las 52 casillas, avanzando según el número de caras marcadas que se obtienen al lanzar los 5 colorines.

Durante la Edad media los juegos con dados se extendieron por toda Europa y se fue adquiriendo un conocimiento empírico acerca de ellos. Fue en esa época cuando se acuñó el término azar, el cual es de origen árabe y surgió a fines del siglo XI. En una de las caras de los dados que se utilizaban para jugar, estaba dibujada una flor, representando un resultado desfavorable para quien lo obtenía. La expresión árabe es az-zahr, la cual significa la flor. Por extensión se llamaba también az-zahr al dado y también se entendía por az-zahr el lanzar los dados.

La experiencia fue indicando que algunas combinaciones que se obtenían al lanzar los dados eran menos frecuentes que otras y quienes jugaban las conocían. Sin embargo, hasta la Edad Media, nadie intentó desarrollar una teoría matemática de los juegos con dados. No fue sino hasta el Renacimiento cuando se comenzó a teorizar acerca de ese tema.

El primer estudio sistemático de problemas relacionados con el lanzamiento de varios dados y los diferentes resultados que pueden obtenerse lo realizó Gerolamo Cardano (1501-1576), en su libro *Liber de ludo aleae* (Libro de los juegos de azar), sin embargo su trabajo tuvo poca influencia, pues si bien fue escrito en 1564, se publicó en 1663. Para ese entonces Blaise Pascal (1623-1662), Pierre de Fermat (1601-1665) y Christiaan Huygens (1629-1695) ([**33**], [**49**]) habían ya resuelto algunos problemas de probabilidad, los cuales se consideran como los que sirvieron de base para desarrollar el Cálculo de Probabilidades.

Uno de los primeros problemas matemáticos que se planteó al considerar los juegos con dados fue el de determinar cuántos resultados distintos pueden obtenerse al lanzar n dados. La primera solución correcta conocida de este problema se encuentra en un poema titulado "De Vetula" y escrito por Richard de Fournival (1200-1250). Ahí se afirma que 3 dados pueden caer en un total de 216 caminos ([**65**]).

La primera referencia conocida a una relación entre las diferentes posibilidades de ocurrencia de un evento y la frecuencia con que éste se observa, se encuentra en los comentarios a una publicación de "La Divina Comedia" que en el año 1477 hizo Benvenuto d'Imola (1320-1388). Dice ahí: "Concerniente a estos lanzamientos (de dados) debe observarse que los dados son cuadrados y cualquier cara puede caer, así que un número que pueda aparecer en más caminos debe ocurrir más frecuentemente, como en el siguiente ejemplo: con tres dados, tres es el más pequeño número que puede obtenerse y sólo se obtiene con tres ases; cuatro puede obtenerse sólo en un camino, con un dos y dos ases" ([**65**]).

Fue en el año 1654 cuando se plantearon algunos problemas cuyas soluciones condujeron al establecimiento de reglas generales para calcular probabilidades. Pareciera, por la referencias que hay, que en esa época los juegos con dados eran muy populares y se conocían algunas reglas que permitían a los jugadores conocer, de manera aproximada y siempre con incertidumbre, que posibilidades tenían de ganar al realizar una apuesta. Antoine Gombaud (1607-1684), conocido como el chevalier de Méré, jugó un papel importante para que se desarrollara un Cálculo de Probabilidades, no por aportaciones que hubiera hecho, sino porque fue él quien, en el año 1654, le planteó a Pascal dos problemas que, al ser resueltos de manera independiente por Pascal y Fermat en el mismo año, y por Huygens 3 años más tarde, marcaron la pauta para poder dar solución a una diversidad de problemas de probabilidad, lo cual condujo a una formulación general de las reglas que se requerían para el desarrollo de un Cálculo de Probabilidades.

Era del conocimiento del Chevalier de Méré que al lanzar un dado 4 veces consecutivas, había más posibilidades de obtener el número 6 por lo menos en uno de los lanzamientos que no obtener ninguno: sin embargo tenía duda acerca de cuántos lanzamientos de un par de dados se requieren para que haya más posibilidades de obtener par de seises por lo menos en uno de los lanzamientos del par de dados, que no obtener ninguno. Pensaba que la solución era tal vez 24 lanzamientos ya que argumentaba: Al lanzar un dado se tienen 6 posibles resultados y en ese caso se requieren 4 lanzamientos para apostar, con ventaja, que se obtiene el número 6 por lo menos en una ocasión; al lanzar un par de dados se tienen 36 posibles resultados y como 6 es a 4 como 36 a 24, entonces se requieren 24 lanzamientos del par de dados para apostar, con ventaja, que se obtiene par de seises por lo menos en una ocasión.

Al tener duda acerca de este resultado, le planteó el problema a Pascal, quien encontró que se requieren no 24, sino 25 lanzamientos del par de dados para que sea más favorable obtener por lo menos un par de seises que no obtener ninguno. Pascal a su vez le planteó el problema a Fermat, quien, sin conocer el método empleado por Pascal, llegó a la misma conclusión.

Se desconocen los métodos que utilizaron Pascal y Fermat para resolver el problema de los dados planteado por de Mére; sin embargo, en una carta dirigida a Pascal, Fermat desarrolló la solución de un problema con dados y de ahí podría inferirse que el método utilizado por Fermat para resolver el problema del lanzamiento de un par de dados fue como sigue:

Si trato de hacer un par de seises en n lanzamientos de un par de dados y, después que el dinero está en juego, convenimos en que no haré el primer lanzamiento, entonces es necesario que saque del juego $\frac{1}{36}$ del total, el cual denotaremos por A. Si después de eso, convenimos en que no haré el segundo lanzamiento, entonces debo sacar $\frac{1}{36}$ de lo restante, es decir de $A - \frac{1}{36}A = \frac{35}{36}A$, con lo cual se obtiene $\frac{35}{1296}A$; y si después de eso convenimos en que no haré el tercer lanzamiento, entonces debo sacar $\frac{1}{36}$ de lo restante, es decir de $\frac{35}{36}A - \frac{35}{1296}A = \frac{1225}{1296}A$, con lo cual se obtiene $\frac{1225}{46656}A$; si todavía se conviene en que no haga el cuarto lanzamiento debo sacar $\frac{1}{36}$ de lo restante, es decir de $\frac{1225}{1296}A - \frac{1225}{46656}A = \frac{42875}{46656}A$, con lo cual se obtiene $\frac{42875}{1679616}A$. Este proceso continuaría hasta que la suma de lo que le corresponde al jugador, desde el primer paso hasta el último, sea mayor que $\frac{1}{2}A$.

Como puede verse, así planteada, la solución de Fermat parece demasiado laboriosa, con números demasiado grandes; sin embargo, se puede simplificar el proceso de una manera que parece obvia, pero se desconoce si Fermat lo hizo. En efecto, el razonamiento anterior se ´puede escribir de la siguiente manera:

Si trato de hacer un par de seises en n lanzamientos de un par de dados y, después que el dinero está en juego, convenimos en que no haré el primer lanzamiento, entonces es necesario que saque del juego $\frac{1}{36}$ del total, el cual denotaremos por A. Si después de eso, convenimos en que no haré el segundo lanzamiento, entonces debo sacar $\frac{1}{36}$ de lo restante, es decir de $A - \frac{1}{36}A = \frac{35}{36}A$, con lo cual se obtiene $\frac{35}{(36)^2}A$; y si después de eso convenimos en que no haré el tercer lanzamiento, entonces debo sacar $\frac{1}{36}$ de lo restante, es decir de $\frac{35}{36}A - \frac{35}{(36)^2}A = \frac{(35)^2}{(36)^2}A$, con lo cual se obtiene $\frac{(35)^2}{(36)^3}A$. si todavía se conviene en que no haga el cuarto lanzamiento debo sacar $\frac{1}{36}$ de lo restante, es decir de $\frac{(35)^2}{(36)^2}A - \frac{(35)^2}{(36)^3}A = \frac{(35)^3}{(36)^3}A$, con lo cual se obtiene $\frac{(35)^3}{(36)^4}A$. De aquí ya puede verse que, en el n-simo paso, le corresponde al jugador $\frac{(35)^{n-1}}{(36)^n}$. Por lo tanto, la no realización de los n lanzamientos le vale al jugador la suma:

$$\frac{1}{36}A + \frac{35}{(36)^2}A + \frac{(35)^2}{(36)^3}A + \frac{(35)^3}{(36)^4}A + \cdots + \frac{(35)^{n-1}}{(36)^n}A = \frac{\frac{1}{36} - \frac{(35)^n}{(36)^{n+1}}}{1 - \frac{35}{36}}A = \left[1 - \left(\frac{35}{36}\right)^n\right]A$$

Así que se trata de obtener el más pequeño valor de n para el cual $1 - \left(\frac{35}{36}\right)^n$ sea mayor que $\frac{1}{2}$, cuya solución es $n = 25$.

Las soluciones de Huygens a los problemas resueltos por Pascal y Fermat las expuso en su libro *De ratiociniis in ludo aleae* (Del razonamiento en los juegos de azar), publicado en el año 1657. En ese libro encontramos la primera teorización del Cálculo de Probabilidades.

Huygens tomó como punto de partida la siguiente hipótesis, la cual es básicamente una definición:

> En un juego, la posibilidad que se tiene de ganar alguna cosa tiene un valor tal que, si se pose ese valor, se puede uno procurar la misma posibilidad en un juego equitativo.

Por un juego equitativo, Huygens entendía un juego que no va en detrimento de ninguno de los jugadores. Incluye el caso de un juego entre un número cualquiera de jugadores en el cual, o bien todos los jugadores tienen la misma posibilidad de ganar cierta cantidad, o bien cada uno de los jugadores tiene la misma posibilidad de ganar cierta cantidad que de perderla.

De su hipótesis, Huygens dedujo tres proposiciones, de las cuales, las dos primeras son un caso particular de la siguiente:

PROPOSICIÓN 12.1. *Tener iguales posibilidades de obtener a_1, a_2, a_3, ..., a_n tiene un valor de $\frac{a_1+a_2+a_3+\cdots+a_n}{n}$.*

Demostración

Llamemos P a quien tiene iguales posibilidades de obtener a_1, a_2, a_3, ..., a_n. Consideremos entonces un juego entre P y otros $n-1$ jugadores, que llamaremos Q_2, Q_3, Q_4, ..., Q_n, en el cual los n tienen las mismas posibilidades de ganar y cada uno apuesta $\frac{a_1+a_2+a_3+\cdots+a_n}{n}$. Quien gane se lleva todas las apuestas, es decir, $a_1+a_2+a_3+\cdots+a_n$. Evidentemente, éste es un juego equitativo. Si, además, para cada $k \in \{2, 3, \ldots, n\}$, P acuerda con Q_k que, si alguno de los dos gana el juego, el ganador le dará al otro la cantidad a_k. Cada uno de estos acuerdos es equitativo y no va en detrimento de ninguno de los otros jugadores; así que el juego continúa siendo equitativo. Si P gana el juego, una vez que entrega lo acordado con cada uno de los otros jugadores, obtiene a_1.Si el juego lo gana Q_k, P obtiene a_k. Así que P tiene iguales posibilidades de obtener a_1, a_2, a_3, ..., a_n, en un juego equitativo.

■

PROPOSICIÓN 12.2. *Tener r posibilidades de obtener a y s posibilidades de obtener b, las posibilidades siendo equivalentes, tiene un valor de $\frac{ra+sb}{r+s}$.*

Demostración

Llamemos P a quien tiene r posibilidades de obtener a y s posibilidades de obtener b. Consideremos entonces un juego entre P y otros $r+s-1$ jugadores, en el cual los $r+s$ tienen las mismas posibilidades de ganar y cada uno apuesta $\frac{ra+sb}{r+s}$. Quien gane se lleva todas las apuestas, es decir, $ra+sb$. Evidentemente, éste es un juego equitativo. Supongamos, además, que P acuerda con cada uno de s jugadores que, si alguno de los dos gana el juego, el ganador le dará al otro la cantidad b; llamaremos a este conjunto de s jugadores, el grupo 1. Finalmente, supongamos también que, con cada uno del resto de los jugadores, P acuerda que, si alguno de los dos gana el juego, el ganador le dará al otro la cantidad a; llamaremos a este conjunto de $r-1$ jugadores, el grupo 2. Cada uno de estos acuerdos es equitativo y no va en detrimento de ninguno de los otros jugadores; así que el juego continúa siendo equitativo. Si P gana el juego, una vez que entrega lo acordado con cada uno de los otros jugadores, obtiene a. Si el juego lo gana algún jugador del grupo 1, P obtiene b, mientras que si lo gana alguno del grupo 2, obtiene a. Como los $r+s$ jugadores tienen las mismas posibilidades de ganar, P tiene entonces r posibilidades de obtener a y s posibilidades de obtener b, en un juego equitativo.

■

Podemos ver que la hipótesis de Huygens es básicamente la definición del concepto de esperanza de lo que se obtiene sobre lo que está en juego; es decir, la esperanza de una variable aleatoria definida como lo que obtiene el jugador.

La solución que dio Huygens al problema planteado por el Chevalier de Méré, concerniente al lanzamiento de un par de dados, es la siguiente:

Llamemos A a la cantidad que está en juego. Quien juega a un solo lanzamiento tiene 1 posibilidad de obtener A y 35 posibilidades de no obtener nada, así que el valor que tiene este juego para ese jugador es $\frac{1}{36}A$.

Quien juega a dos lanzamientos, en su primer lanzamiento tiene 1 posibilidad de obtener A y 35 posibilidades de obtener $\frac{1}{36}A$ (por el primer paso), así que el valor que tiene este juego para ese jugador es $\frac{A+35\left(\frac{1}{36}A\right)}{36} = \frac{71}{(36)^2}A$.

Quien a cuatro lanzamientos, obtiene A si sale par de seises en alguno de los primeros dos lanzamientos, si no, por el segundo paso, obtiene $\frac{71}{(36)^2}A$; pero, también por el segundo paso, hay 71 posibilidades de obtener par de seises en alguno de los dos primeros lanzamientos y $(36)^2 - 71 = 1225$ posibilidades de no obtenerlo; por lo tanto, el valor que tiene este juego para ese jugador es $\frac{71A+1225\left(\frac{71}{(36)^2}A\right)}{(36)^2} = \frac{178991}{(36)^4}A$.

Quien a ocho lanzamientos, obtiene A si obtiene par de seises en alguno de los primeros cuatro lanzamientos, si no, por el paso anterior, obtiene $\frac{178991}{(36)^4}A$; pero, también por el paso anterior, hay 178991 posibilidades de obtener par de seises en alguno de los cuatro primeros lanzamientos y $(36)^4 - 178991 = 1500625$ posibilidades de no obtenerlo; por lo tanto, el valor que tiene este juego para ese jugador es $\frac{178991A+1500625\left(\frac{178991}{(36)^4}A\right)}{(36)^4} = \frac{569234516831}{(36)^8}A$.

Quien juega a 16 lanzamientos, obtiene A si sale par de seises en alguno de los primeros ocho lanzamientos; si no, por el paso anterior, obtiene $\frac{569234516831}{(36)^8}A$; ; pero, también por el paso anterior, hay 569234516831 posibilidades de obtener par de seises en alguno de los ocho primeros lanzamientos y

$$(36)^8 - 569234516831 = 2251875390625$$

posibilidades de no obtenerlo; por lo tanto, el valor que tiene este juego para ese jugador es igual a

$$\frac{569234516831A + 2251875390625\left(\frac{569234516831}{(36)^8}A\right)}{(36)^8} = \frac{28877183350439045465 01311}{(36)^{16}}A.$$

Quien juega a 24 lanzamientos, obtiene A si sale par de seises en alguno de los primeros ocho lanzamientos; si no, por el paso anterior, obtiene $\frac{2887718335043904546501311}{(36)^{16}}A$;

pero, por el paso previo al anterior, hay 569234516831 posibilidades de obtener par de seises en alguno de los ocho primeros lanzamientos y

$$(36)^8 - 569234516831 = 2251875390625$$

posibilidades de no obtenerlo; por lo tanto, el valor que tiene este juego para ese jugador es igual a

$$\frac{569234516831A + 2251875390625\left(\frac{2887718335043904546501311}{(36)^{16}}A\right)}{(36)^8} = \frac{11033126465283976852912127963392284191}{(36)^{24}}A.$$

Quien juega a 25 lanzamientos, obtiene A si sale par de seises en su primer lanzamiento; si no, por el paso anterior, obtiene

$$\frac{11033126465283976852912127963392284191}{(36)^{24}}A;$$

por lo tanto, el valor que tiene este juego para ese jugador es igual a

$$\frac{A + 35\left(\frac{11033126465283976852912127963392284191}{(36)^{24}}A\right)}{36} = \frac{408611683992293747092011689842522621501}{(36)^{25}}A.$$

Ahora bien:

$$\frac{11033126465283976852912127963392284191}{(36)^{24}} \approx 0.4914$$

$$\frac{408611683992293747092011689842522621501}{(36)^{25}} \approx 0.5055$$

Por lo tanto, quien juega a 24 lanzamientos tiene más posibilidades de perder que de ganar, mientras que quien juega a 25 lanzamientos tiene más posibilidades de ganar que de perder.

Huygens resolvió el problema del lanzamiento de un par de dados como se describió arriba, con los números que aparecen en los numeradores, sin buscar alguna simplificación; incluso, únicamente escribe los cocientes hasta el caso de quien juega a 4 lanzamientos y después sólo describe los pasos siguientes; pero llega a la misma conclusión que Pascal y Fermat; a saber, que se requieren por lo menos 25 lanzamientos para tener más posibilidades de ganar que de perder.

Más allá de los cálculos que realiza Huygens, los cuales pueden simplificarse, podemos observar que en sus soluciones utiliza valores condicionales de un juego, lo cual, en terminología moderna, corresponden a esperanzas condicionales.

El otro problema que el Chevalier de Méré planteó a Pascal es el siguiente:

¿Cómo deben repartirse las apuestas en un juego que se interrumpe? Por ejemplo, suponiendo que dos jugadores, A y B, apuestan 32 pesos cada uno en un juego que consiste de partidas consecutivas, en cada una de las cuales cada jugador tiene la misma posibilidad de ganarla, de tal manera que quien gane una partida acumula un punto y el juego es ganado por quien obtenga primero cuatro puntos, ¿cómo deben de repartirse las apuestas en caso de que el juego se interrumpa cuando el jugador A ha ganado dos puntos y B un punto?

Este problema fue el que más interés provocó debido a que pocos lograron encontrar la solución correcta. La solución que dieron Pascal, Fermat y Huygens consistió en encontrar la probabilidad que cada jugador tiene de ganar el juego, partiendo de la situación en la cual se interrumpe. La solución de Fermat fue básicamente la siguiente:

Al jugador A le faltan 2 partidas para ganar y al jugador B, 3 partidas, entonces el juego termina en a lo más 4 partidas adicionales. Denotando por la letra a el que A gane una partida y por la letra b el que gane B, los posibles resultados de 4 partidas son los siguientes:

$(a, a, a, a), (a, a, a, b), (a, a, b, a), (a, a, b, b), (a, b, a, a), (a, b, a, b), (a, b, b, a), (b, a, a, a),$

$(b, a, a, b), (b, a, b, a), (b, b, a, a), (b, b, b, b), (b, b, b, a), (b, b, a, b), (a, b, b, b), (b, a, b, b)$

donde, por ejemplo, (b, b, a, b) significa que A gana sólo la tercera partida y B las otras 3.

De estos 16 posibles resultados, hay 11 que hacen ganar al jugador A, a saber, (a, a, a, a), (a, a, a, b), (a, a, b, a), (a, a, b, b), (a, b, a, a), (a, b, a, b), (a, b, b, a), (b, a, a, a), (b, a, a, b), (b, a, b, a) y (b, b, a, a). Los 5 restantes, (b, b, b, b), (b, b, b, a), (b, b, a, b), (a, b, b, b) y (b, a, b, b), hacen ganar al jugador B. Por lo tanto, las apuestas se deben repartir en la proporción $11 : 5$.

Podemos observar que en la solución de Fermat parece haber un problema, pues, para contar los casos en que gana cada jugador, considera que se juegan las 4 partidas y en algunos de esos casos el juego se termina antes de llegar a la cuarta. En realidad eso no representa problema ya que, por ejemplo, si A gana las dos primeras partidas (de las 4), en cuyo caso se acabaría ahí el juego, lo que hace Fermat es descomponer ese caso, es decir (a, a) en 4 casos, a saber, (a, a, a, a), (a, a, a, b), (a, a, b, a) y (a, a, b, b); el caso (a, a) tiene 1 posibilidad de 4; los cuatro casos juntos (a, a, a, a), (a, a, a, b), (a, a, b, a) y (a, b, a, a) tienen 4 posibilidades de 16; así que las proporciones son ambas iguales a $\frac{1}{4}$. Algo similar puede argumentarse en las otras situaciones. El caso (a, b, a) se descompone en (a, b, a, a) y (a, b, a, b); el caso (b, a, a) se descompone en (b, a, a, a) y (b, a, a, b); el caso (b, b, b) se descompone en (b, b, b, b)

y (b, b, b, a). El objetivo de esas descomposiciones es, como lo decía Fermat, "hacer todos los azares iguales", es decir, en lenguaje moderno, todos los casos igualmente probables.

La solución de Fermat al problema de la división de apuestas deja ver que Fermat utilizaba ya lo que más tarde se llamaría la definición clásica de probabilidad, la cual se puede aplicar siempre que los posibles resultados se puedan considerar equiprobables:

$$P(A) = \frac{\#\text{ de posibles resultados que producen la ocurrencia de } A}{\#\text{ total de posibles resultados}}$$

12.2. Jacques Bernoulli

En el año 1713 se publicó un libro de Jacques Bernoulli (1654-1705), titulado *Ars Conjectandi* (El arte de conjeturar) ([**4**]). En ese libro, Bernoulli sentó las bases para el desarrollo posterior del Cálculo de Probabilidades. Formuló métodos generales para resolver problemas de probabilidad, dando así una formulación teórica más sólida que la de Huygens. Esto además de enfrentar el problema de probar que la frecuencia relativa con la que se presenta un evento se aproxima a la probabilidad del mismo a medida que el número de observaciones se hace más grande; en esta búsqueda demostró el primero de los teoremas límite del Cálculo de Probabilidades, el cual daría la pauta para una investigación que se prolongó por más de 200 años, hasta llegar a una formulación general que culminaría hacia el año 1930.

Bernoulli comenzó su libro con un análisis de los problemas resueltos por Huygens en su obra *De ratiociniis in ludo aleae*. Para esto tomó como base la hipótesis de Huygens, a la cual consideró como el **principio fundamental del arte de conjeturar**. Con respecto al problema planteado por el Chevalier de Méré, concerniente al lanzamiento de un par de dados, Bernoulli planteó el uso de letras, en lugar de números, con el objeto de obtener una solución general de ese problema:

Consideremos el lanzamiento de uno o muchos dados, llamemos b al número de casos en los cuales, en cada lanzamiento, se obtiene éxito en lo que se propone, c al número de casos en los cuales no se obtiene éxito y $a = b + c$. Si P juega a ganar en un lanzamiento, tiene $a - c$ casos en los cuales tiene éxito, obteniendo entonces el total en juego, el cual considera igual a 1 para simplificar; en cualquiera de los otros c casos, no obtiene nada, de manera que el valor del juego para P es igual a $\frac{a-c}{a}$. Si juega a dos lanzamientos, en el primero tiene $a - c$ casos en los cuales obtiene $1 = \frac{a}{a}$ y c casos en los cuales regresa a la situación precedente; así que el valor del juego para P es igual a $\frac{(a-c)\frac{a}{a} + c\frac{a-c}{a}}{a} = \frac{a^2-c^2}{a^2}$. Si juega a tres lanzamientos, en el primero tiene $a - c$ casos en los cuales obtiene $1 = \frac{a^2}{a^2}$ y c casos en los cuales regresa a la situación precedente; así que el valor del juego para P es igual a $\frac{(a-c)\frac{a^2}{a^2} + c\frac{a^2-c^2}{a^2}}{a} = \frac{a^3-c^3}{a^3}$. En general, si juega a n lanzamientos, el valor del juego para P es igual a $\frac{a^n-c^n}{a^n}$.

Bernoulli agregó otro método para resolver este problema: Si P juega a obtener éxito en el primer lanzamiento, el valor del juego es igual a $\frac{a-c}{a} = \frac{b}{a}$. Si juega a obtener éxito en el segundo lanzamiento, fallando en el primero, tiene c casos en los cuales regresa a la situación del inicio; así que el valor del juego es igual a $\frac{c\frac{b}{a}}{a} = \frac{bc}{a^2}$. Si juega a obtener éxito en el tercer lanzamiento, fallando en los dos primeros, tiene c casos en los cuales regresa a la situación precedente; así que el valor del juego es igual a $\frac{c\frac{bc}{a^2}}{a} = \frac{bc^2}{a^3}$. En general, si juega a obtener éxito en el k-ésimo lanzamiento, fallando en los $k-1$ primeros, el valor del juego es igual a $\frac{c\frac{bc}{a^2}}{a} = \frac{bc^{k-1}}{a^k}$. Por lo tanto, si juega a obtener éxito en por lo menos uno de n lanzamientos, el valor del juego es la suma de los n casos particulares descritos antes; es decir:

$$\frac{b}{a} + \frac{bc}{a^2} + \frac{bc^2}{a^3} + \cdots + \frac{bc^{n-1}}{a^n} = \frac{\frac{b}{a} - \frac{bc^{n-1}}{a^n}\frac{c}{a}}{1-\frac{c}{a}} = \frac{a^n - c^n}{a^n}$$

Bernoulli agregó un tercer método: El lanzamiento de un dado n veces es equivalente a lanzar una vez n dados. Consideremos entonces n dados, cada uno con a caras, de las cuales hay c en las que no se obtiene éxito. El número de casos que se pueden obtener al lanzar los n dados es igual a a^n, de los cuales hay c^n casos en que no se obtiene éxito con ninguno de ellos; así que hay $a^n - c^n$ casos en los cuales se obtiene éxito por lo menos con uno de los dados, en cuyo caso P obtiene 1, y c^n casos en los cuales obtiene 0; por lo tanto, el valor del juego es igual a $\frac{a^n - c^n}{a^n}$.

En el tercer método de Bernoulli, aunque, al igual que Huygens, utiliza esperanzas, se puede ver claramente la definición clásica de probabilidad, además de que resuelve el problema de la manera más simple, razonando sobre los casos en que no se obtiene éxito. Este hecho muestra que no siempre la solución más simple es la primera que se ocurre e incluso puede no ser evidente; ni Pascal, ni Fermat, ni Huygens encontraron esta forma simple de resolver el problema planteado. Lo inmediato o simple de una solución a un problema requiere, a veces, de ensayos de solución y de maduración de determinados conceptos.

En el segundo método vemos el uso de la propiedad de la aditividad finita de la función de probabilidad.

Huygens consideró otro problema con dados, el cual, al ser generalizado por Bernoulli, adquiriría una importancia central en el desarrollo del Cálculo de Probabilidades:

¿Cuántos dados se requieren lanzar para que sea más favorable obtener por lo menos dos seises?

Con relación a este problema, Bernoulli encontró que la probabilidad de obtener exactamente k seises en n lanzamientos de un dado es igual a $\binom{n}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{n-k}$. Este resultado es de fundamental importancia en su trabajo pues con él se puede calcular la probabilidad de obtener una frecuencia de seises igual a $\frac{k}{n}$ en n lanzamientos de un dado y de aquí encontrar una relación entre la frecuencia de ocurrencia de un evento y su probabilidad, para obtener lo que se llama el Teorema de Bernoulli.

Otro problema de gran importancia que planteó y resolvió Huygens en su libro es el siguiente:

Dos jugadores, P y Q, juegan a lanzar alternadamente un par de dados. El juego comienza lanzando P el par de dados, con la condición de que si obtiene una suma igual a 6, gana el juego; en caso contrario, el juego continúa lanzando Q el par de dados, con la condición de que si obtiene una suma igual a 7, gana el juego; en caso contrario el juego continúa lanzando P el par de dados bajo las condiciones iniciales. ¿Cuáles son las respectivas probabilidades que cada jugador tiene de ganar el juego?

La importancia de ese problema radica en que se refiere a un experimento el cual admite una infinidad de posibles resultados, rebasando el marco de la definición clásica de probabilidad. La solución de Huygens fue como sigue:

Sea x el valor del juego para Q y a el total de las apuestas. El valor del juego para P es entonces $a - x$. Sea además y el valor del juego para Q cuando sea su turno de lanzar los dados. Al iniciarse el juego, Q tiene 5 posibilidades de obtener 0 (cuando P obtiene una suma igual a 6) y 31 posibilidades de obtener y, por lo tanto, $x = \frac{31y}{36}$. Por otra parte, cada vez que Q tenga el turno para lanzar los dados, tiene 6 posibilidades de obtener a y 30 de obtener x, por lo tanto, $y = \frac{6a+30x}{36}$. Resolviendo el sistema de ecuaciones se obtiene $x = \frac{31}{61}a$, de manera que los valores del juego para P y Q, respectivamente, están en la proporción 30 : 31.

Obsérvese que la solución de Huygens se basa en que si P y Q no ganan el juego en su primera oportunidad, se vuelve a la situación inicial.

Bernoulli resolvió este problema estableciendo una progresión geométrica para la probabilidad que cada jugador tiene de ganar el juego:

En lugar de dos jugadores, supongamos que hay una infinidad, cada uno de los cuales tiene una oportunidad para ganar, lanzando consecutivamente el par de dados, con la condición de que el primer jugador de orden impar que obtenga 6 puntos, o el primer jugador de orden par que obtenga 7 puntos, gana el juego. Denotemos por b y c al número de casos favorables y desfavorables, respectivamente, a la obtención de 6 puntos al lanzar el par de dados; por e y f al número de casos favorables y desfavorables, respectivamente, a la obtención de 7 puntos y por a al total de casos. Las posibilidades que tiene cada jugador de ganar el juego están dadas por:

Jugador 1: b casos de un total de a casos.

Jugador 2: ce casos de un total de a^2 casos.

Jugador 3: cfb casos de un total de a^3 casos.

Jugador 4: $cfce$ casos de un total de a^4 casos.

Jugador 5: $cfcfb$ casos de un total de a^5 casos.

$\vdots$

Si suponemos ahora que todos los jugadores de orden impar son sustituidos por P y todos los jugadores de orden par son sustituidos por Q, las suertes de P y Q están dadas, respectivamente, por:

$$\frac{b}{a} + \frac{cfb}{a^3} + \frac{c^2f^2b}{a^5} + \frac{c^3f^3b}{a^7} + \cdots = \frac{\frac{b}{a}}{1-\frac{cf}{a^2}} = \frac{ab}{a^2-cf}$$

$$\frac{ce}{a^2} + \frac{cfce}{a^4} + \frac{c^2f^2ce}{a^6} + \frac{c^3f^3ce}{a^8} + \cdots = \frac{\frac{ce}{a^2}}{1-\frac{cf}{a^2}} = \frac{ce}{a^2-cf}$$

Por lo tanto, la suerte de P es a la de Q como ab es a ce; es decir, como (36) (5) es a (31) (6); o, de manera equivalente, como 30 es a 31.

Bernoulli estaba estableciendo entonces que la probabilidad de que un jugador gane el juego es igual a la suma de las probabilidades de que gane en cada uno de los posibles turnos que tiene, los cuales son una infinidad. En otras palabras, está implícita en el resultado la propiedad de σ- aditividad de la función de probabilidad.

El método de Bernoulli fue retomado un poco más adelante, en el año 1718, por Abraham de Moivre en su libro ([**29**]), sin embargo, aunque aparentemente era conocido, no se utilizó durante el resto del siglo XVIII y todo el XIX, de manera que la propiedad de σ-aditividad de la función de probabilidad quedó relegada en la sistematización de Laplace, la cual perduró hasta principios del siglo XX.

El poco interés que atrajo el método de Bernoulli puede no haber sido circunstancial, sino que parece obedecer a la concepción de la probabilidad que está implícita en su formulación clásica, la cual está basada en la equiprobabilidad de los diferentes resultados de un experimento aleatorio, cuyo número debe entonces ser finito, de manera que los problemas donde se realizan repeticiones indefinidas de experimentos aleatorios únicamente pueden tratarse mediante aproximaciones a través de sus correspondientes límites. Los problemas considerados arriba, donde el número de casos posibles es infinito, sin ser de probabilidades continuas, caen dentro de esta categoría de problemas, que no rebasan el marco clásico. En efecto, la solución de Bernoulli al problema planteado por Huygens, por ejemplo, puede plantearse como una distribución límite considerando las probabilidades que cada jugador tiene de ganar en los primeros $2n$ lanzamientos y haciendo tender luego n a ∞. Si llamamos $P_n(A)$ y $P_n(B)$ a estas probabilidades se obtiene, utilizando la notación mencionada más arriba:

$$P_n(A) = \sum_{k=1}^{n} p(\omega_{2k-1}) = \sum_{k=1}^{n} (\tfrac{31}{36})^{k-1} (\tfrac{30}{36})^{k-1} \tfrac{5}{36} = \tfrac{30}{61} \left[1 - (\tfrac{31}{36})^n (\tfrac{30}{36})^n\right]$$

$$P_n(B) = \sum_{k=1}^{n} p(\omega_{2k}) = \sum_{k=1}^{n} (\tfrac{31}{36})^{k} (\tfrac{30}{36})^{k-1} \tfrac{6}{36} = \tfrac{31}{61} \left[1 - (\tfrac{31}{36})^n (\tfrac{30}{36})^n\right]$$

Veremos más adelante que efectivamente, incluso todavía en los 20's del siglo XX se daba esa interpretación a la solución de Bernoulli.

Las soluciones de Bernoulli a los problemas resueltos por Huygens representaron un avance significativo en el camino de dotar al Cálculo de Probabilidades de una teoría que permitiera ir resolviendo problemas cada vez más complejos. Sin embargo, la aportación central de

Bernoulli la encontramos en la última parte de su libro, donde planteó un problema de singular importancia, el cual sería la base para el desarrollo posterior de la teoría durante un periodo de más de 200 años. Fue a partir de ese resultado que el Cálculo de Probabilidades comenzó a ganarse un lugar importante dentro de la Matemática.

Escribió Bernoulli en su libro:

> *"Parece que, para hacer una hipótesis correcta sobre un hecho cualquiera, sólo es necesario calcular exactamente el número de casos posibles y, entonces, determinar las veces que puede posiblemente ocurrir un caso más que otro. Pero aquí, inmediatamente, surge nuestra mayor dificultad, porque este procedimiento se puede aplicar únicamente a muy pocos fenómenos; de hecho, casi exclusivamente a los relacionados con los juegos de azar ... pero hay otro camino que nos conduce a lo que buscamos, y nos permite, por lo menos, hallar a posteriori lo que no podemos determinar a priori, o sea, averiguando a partir de los resultados observados en numerosos casos similares.*
>
> *Ha de suponerse, a este respecto, que, bajo condiciones similares, la ocurrencia (o no ocurrencia) de un suceso en el futuro seguirá la misma pauta que se ha observado para sucesos iguales en el pasado ... Lo que aún tiene que ser averiguado es si, cuando se aumenta el número de observaciones, también se sigue aumentando la probabilidad de que la proporción registrada de casos favorables y desfavorables se aproxime a la verdadera relación ... Este es el problema que he decidido publicar aquí, después de haber trabajado sobre él durante veinte años." ([**72**])*

Obsérvese que, en su razonamiento, Bernoulli supone que en los fenómenos aleatorios existe una regularidad, a saber, que la frecuencia relativa con la que se observa que ocurre un evento, se mantiene en el futuro como se observó en el pasado. A esta propiedad la llamaremos principio de regularidad de la frecuencia relativa con la que ocurre un evento.

El resultado al que hace referencia Bernoulli en su libro es el ahora llamado teorema de Bernoulli, el cual, utilizando terminología moderna, se puede enunciar como sigue:

> Sea $\mathcal{E}$ un experimento aleatorio que admite t posibles resultados equiprobables y A un evento relativo a ese experimento, para el cual hay r resultados que favorecen su ocurrencia. Consideremos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento $\mathcal{E}$, de tal manera que cada repetición es independiente de las otras. Sea X_{nt} el número de veces que ocurre el evento A en las primeras nt repeticiones del experimento, entonces:
> $\lim_{n\rightsquigarrow\infty} P\left[\left|\frac{X_{nt}}{nt} - \frac{r}{t}\right| > \frac{1}{t}\right] = 0.$

Sin modificar lo esencial del razonamiento de Bernoulli para demostrar este resultado, se puede enunciar de la siguiente manera:

> Sea $\mathcal{E}$ un experimento aleatorio que admite t posibles resultados equiprobables y A un evento relativo a ese experimento, para el cual hay r resultados que favorecen su ocurrencia. Consideremos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento $\mathcal{E}$, de tal manera que cada repetición es independiente de las otras. Sean X_n el número de veces que ocurre el evento A en las primeras n repeticiones del experimento y ε un número positivo arbitrario, entonces:
> $\lim_{n \rightsquigarrow \infty} P\left[\left|\frac{X_n}{n} - \frac{r}{t}\right| > \varepsilon\right] = 0.$

El resultado de Bernoulli hizo patente que en el modelo teórico que se estaba desarrollando se da efectivamente una correspondencia entre las probabilidades y las frecuencias con que se observan los posibles resultados de un suceso azaroso. Este resultado y otros del mismo tipo que le siguieron sentaron las bases teóricas para aplicar el Cálculo de Probabilidades al estudio de datos estadísticos. Muy pronto esta teoría comenzó a aplicarse al tratamiento de datos como los acumulados en tablas de mortalidad y natalidad.

La idea de la demostración de Bernoulli es la siguiente:

En primer lugar demostró que, para cualquier $k \in \{0, 1, 2, \ldots, n\}$, se tiene:

$P[X_n = k] = \binom{n}{k}\frac{r^k s^{n-k}}{t^n} = \binom{n}{k}\left(\frac{r}{t}\right)^k\left(\frac{s}{t}\right)^{n-k}.$

Después demostró que los términos $t_k = \binom{n}{k}\left(\frac{r}{t}\right)^k\left(\frac{s}{t}\right)^{n-k}$ tienen la propiedad de que crecen con k hasta alcanzar su máximo valor cuando $(n+1)\frac{r}{t} - 1 \leq k \leq (n+1)\frac{r}{t}$ (si $(n+1)\frac{r}{t}$ es un número entero, el valor máximo se alcanza en dos valores de k), después de lo cual decrecen con k.

Finalmente, demostró que, para cualquier $\varepsilon > 0$, se tiene la siguiente relación:

$$\lim_{n \rightsquigarrow \infty} \frac{\sum_{\{k \in \mathbb{N}: k \in [n\frac{r}{t} - n\varepsilon, n\frac{r}{t} + n\varepsilon]\}} \binom{n}{k}\left(\frac{r}{t}\right)^k\left(\frac{s}{t}\right)^{n-k}}{\sum_{\{k \in \mathbb{N}: k \in [0, n\frac{r}{t} - n\varepsilon) \cup (n\frac{r}{t} + n\varepsilon, n]\}} \binom{n}{k}\left(\frac{r}{t}\right)^k\left(\frac{s}{t}\right)^{n-k}} = \infty$$

Es decir, la suma de los términos que se encuentran alrededor del término máximo (en un intervalo de radio aproximadamente igual a $n\varepsilon$) es mucho mayor que la suma de los términos que están fuera de ese rango, a tal grado que esta última suma es despreciable. Esto se puede ver claramente con un ejemplo:

Consideremos el caso en que $t = 20$ y $r = 5$.

Tomando $\varepsilon = \frac{1}{20}$, estaremos cerca del límite de la sumatoria tomando $n = 1000$, en cuyo caso el término máximo se obtiene cuando $k = 250$, y se tiene:

$P\left[\left|\frac{X_n}{n} - \frac{r}{t}\right| \leq \varepsilon\right] = \sum_{\{k \in \mathbb{N}: k \in [250-1000\varepsilon, 250+1000\varepsilon]\}} \binom{n}{k} \left(\frac{r}{t}\right)^k \left(\frac{s}{t}\right)^{n-k}$

$= \sum_{k=200}^{300} \binom{1000}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{1000-k} = 0.99977.$

Tomando $\varepsilon = \frac{1}{100}$, estaremos cerca del límite de la sumatoria tomando $n = 10000$, en cuyo caso el término máximo se obtiene cuando $k = 2500$, y se tiene:

$P\left[\left|\frac{X_n}{n} - p\right| \leq \varepsilon\right] = \sum_{\{k \in \mathbb{N}: k \in [2500-10000\varepsilon, 2500+10000\varepsilon]\}} \binom{n}{k} \left(\frac{r}{t}\right)^k \left(\frac{s}{t}\right)^{n-k}$

$= \sum_{k=2400}^{2600} \binom{10000}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{10000-k} = 0.979\,72.$

En la siguiente figura se encuentran graficados los valores de los términos t_k para el caso $\varepsilon = \frac{1}{20}$ y $n = 1000$.

$n = 1000$, $p = \frac{1}{4}$, $\varepsilon = \frac{1}{20}$

Así que, como puede verse, el teorema de Bernoulli es un resultado de Cálculo Combinatorio.

Escrito en términos de probabilidades, el resultado de Bernoulli se puede escribir como sigue:

$$\lim_{n \rightsquigarrow \infty} \frac{P\left[\left|\frac{X_n}{n} - \frac{r}{t}\right| \leq \varepsilon\right]}{P\left[\left|\frac{X_n}{n} - \frac{r}{t}\right| > \varepsilon\right]} = \infty$$

Considerando que la suma del numerador y el denominador de la expresión anterior es igual a 1, se sigue que:

$$\lim_{n \rightsquigarrow \infty} P\left[\left|\frac{X_n}{n} - \frac{r}{t}\right| > \varepsilon\right] = 0$$

12.3. Teorema de de Moivre-Laplace

La publicación del teorema de Bernoulli hizo renacer el interés por el Cálculo de Probabilidades, el cual, después de la publicación del trabajo de Christiaan Huygens, había quedado relegado, siendo visto únicamente como una curiosidad que tenía que ver exclusivamente con los juegos de azar.

En la búsqueda de mejorar el resultado de Bernoulli, Abraham de Moivre (1667-1754), demostró en 1733 un resultado que también sería de gran importancia en el desarrollo del Cálculo de Probabilidades. En ese año se publicó su artículo titulado *Approximatio ad Summam Terminorum Binomii* $(a+b)^n$ *in Seriem expansi* ([**28**]), en el cual expone un resultado que conduciría a lo que ahora se conoce como el Teorema Central del Límite. El artículo fue publicado en latín y circuló en forma privada. En el año 1738 ese artículo fue incluido en la segunda edición de su libro *The Doctrine of Chances* ([**29**]) con el título *A Method of approximating the Sum of the Terms of the Binomial* $(a+b)^n$ *expanded into a Series, from whence are deduced some practical Rules to estimate the Degree of Assent which is to be given to Experiments*. Con terminología y notación moderna, el resultado de de Moivre puede enunciarse de la siguiente manera ([**42**]):

> **Teorema de de Moivre**. Si d es un número real positivo del orden de $\sqrt{n}$ y, para cada $n \in \mathbb{N}$, X_n es una variable aleatoria con distribución binomial de parámetros n y $p = \frac{1}{2}$, entonces, para n grande, se tiene:
> $P\left[-d \leq X_n - \frac{1}{2}n \leq d\right] \approx \frac{4}{\sqrt{2\pi}} \int_0^{\frac{d}{\sqrt{n}}} e^{-2y^2} dy.$
>
> Tomando $d = \frac{1}{2}x\sqrt{n}$, donde x es un número real positivo, la aproximación de de Moivre toma la siguiente forma:
> $P\left[-x \leq \frac{X_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \leq x\right] \approx \frac{4}{\sqrt{2\pi}} \int_0^{\frac{1}{2}x} e^{-2y^2} dy = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}z^2} dz.$
> Así que podemos expresar el resultado como sigue:
> $\lim_{n \rightsquigarrow \infty} P\left[-x \leq \frac{X_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \leq x\right] = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}z^2} dz.$

Mencionaba de Moivre que su resultado se puede generalizar facilmente para cualquier valor de p. Esta generalización fue expuesta por Pierre Simon Laplace (1749-1827) en su libro *Théorie Analytique des Probabilités*, publicado en el año 1812 ([**54**], [**55**]). En notación moderna, el resultado de Laplace puede escribirse de la siguiente manera ([**42**]):

> **Teorema de de Moivre-Laplace**. Si d es un número real positivo del orden de $\sqrt{n}$ y, para cada $n \in \mathbb{N}$, X_n es una variable aleatoria con distribución binomial de parámetros n y p, entonces, para n grande, se tiene:
> $P\left[-d \leq X_n - np \leq d\right] \approx \frac{2}{\sqrt{\pi}} \int_0^{\frac{d}{\sqrt{2np(1-p)}}} e^{-y^2} dy.$

Tomando $d = x\sqrt{np\,(1-p)}$, donde x es un número real positivo, la aproximación de Laplace toma la siguiente forma:
$P\left[-x \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq x\right] \approx \frac{2}{\sqrt{\pi}} \int_0^{\frac{x}{\sqrt{2}}} e^{-y^2} dy = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}z^2} dz.$
Así que podemos expresar el resultado como sigue:
$\text{lím}_{n \rightsquigarrow \infty} P\left[-x \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq x\right] = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}z^2} dz.$

En su libro Laplace realizó una síntesis del estado del Cálculo de Probabilidades en su época, agregando sus aportaciones. Expuso ahí de manera explícita la definición clásica de probabilidad:

> *"Se ha visto en la introducción que la probabilidad de un evento es el cociente del número de casos que le son favorables entre el número de todos los casos posibles, cuando nada hace pensar que alguno de esos casos debe ocurrir en lugar de los otros, lo cual los hace, para nosotros, igualmente posibles. La justa apreciación de esos casos diversos es uno de los puntos más delicados del Análisis de los azares."*

En seguida enunció y demostró lo que ahora se denomina la propiedad de la aditividad finita de la función de probabilidad:

> *"Si todos los casos no son igualmente posibles, se determinará sus posibilidades respectivas, y entonces la probabilidad del evento será la suma de las probabilidades de cada caso favorable."*

Después consideró el caso en que se tienen varios eventos independientes y mostró que la probabilidad de ocurrencia de todos ellos juntos es igual al producto de sus probabilidades.

Finalmente, en cuanto al cálculo de probabilidades de eventos, enunció y demostró lo que se conoce ahora como la regla del producto:

> *"Si los eventos simples están relacionados entre ellos de manera que la suposición de la ocurrencia del primero influye en la probabilidad de ocurrencia del segundo, se tendrá la probabilidad del evento compuesto, determinando primero la probabilidad del primer evento y después la probabilidad de que el segundo ocurra dado que el primer evento ha ocurrido."*

A pesar de este desarrollo, el azar parecía un concepto que en algún momento perdería importancia pues, para los pensadores de la época, sólo era producto de nuestra ignorancia. Laplace mismo formuló esta idea de manera muy clara, en su libro *Ensayo filosófico sobre las probabilidades*, publicado en el año 1814 ([**56**]). Dice ahí que:

> *"Todos los acontecimientos, aun aquellos que por su insignificancia parecen no depender de las grandes leyes de la naturaleza, constituyen una sucesión tan necesaria como las revoluciones del Sol. Ignorando los vínculos que los ligan al sistema entero del universo, se los ha hecho depender de causas finales o del azar, según que ocurrieran y se sucedieran con regularidad o sin orden aparente; pero esas causas imaginarias han retrocedido gradualmente con los límites de nuestros conocimientos y desaparecen por completo frente a la sana filosofía que no ve en ellas más que la expresión de nuestra ignorancia respecto de las verdaderas causas ... una inteligencia que en un determinado instante pudiera conocer todas las fuerzas que impulsan la naturaleza y la respectiva posición de los seres que la componen y que, además tuviera la suficiente amplitud para someter esos datos al análisis, incluiría en una sola fórmula los movimientos de los mayores cuerpos del universo y del más ligero átomo; nada le sería incierto y tanto el pasado como el futuro estarían en su presencia."*

12.4. El Cálculo de Probabilidades durante la segunda mitad del siglo XIX

La teoría matemática de la probabilidad continuó desarrollándose y fue surgiendo un nuevo concepto, de gran importancia, el de variable aleatoria:

> Una **variable aleatoria** es una variable cuyo valor es aleatorio, depende del resultado del experimento aleatorio en consideración.
>
> Lo que interesa calcular de una variable aleatoria X es la probabilidad con la que toma cada uno de sus posibles valores o la probabilidad de que tome valores en un determinado intervalo. A ese conjunto de probabilidades se le llama la distribución de la variable aleatoria.

Dos cantidades de interés para el estudio de una variable aleatoria son su **esperanza** y su **varianza**. La primera expresa el valor teórico del promedio de los valores que toma la variable aleatoria cuando el experimento aleatorio correspondiente se repite muchas veces. La segunda mide la dispersión de los valores que toma la variable aleatoria, es decir, mide el alejamiento de los valores de la variable aleatoria con respecto a su esperanza.

Si una variable aleatoria X toma valores únicamente en un conjunto finito o infinito numerable, su esperanza se define de la siguiente manera:

$E\left[X\right] = \sum_x xP\left[X = x\right]$.

La varianza de X se suele denotar por $\sigma^2\left(X\right)$ y se define de la siguiente manera:

$\sigma^2\left(X\right) = E\left[\left(X - E\left[X\right]\right)^2\right]$.

Continuando con el estudio de los teoremas de Bernoulli y de de Moivre, se obtuvieron generalizaciones de esos resultados. En particular, la "escuela rusa" hizo grandes aportes a partir de la segunda mitad del siglo XIX:

En el año 1867, Pafnuty Lvovich Chebyshev (1821-1894) demostró una forma general del teorema de Bernoulli (**Ley débil de los grandes números**) ([**18**]):

> Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de varianza finita. Entonces, para cualquier $\varepsilon > 0$, se tiene:
> $\lim_{n \rightsquigarrow \infty} P\left[\left|\frac{X_1+\ldots+X_n}{n} - \mu\right| > \varepsilon\right] = 0$,
> donde μ es la esperanza común de $X_1, X_2, \ldots$.

En el año 1900, Aleksandr Mikhailovich Lyapunov (1857-1918) demostró una forma general del teorema de de Moivre (**Teorema Central del Límite**) ([**63**]):

> Si $X_1, X_2, \ldots$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (con tercer momento finito), entonces:
> $\lim_{n \rightsquigarrow \infty} P\left[a < \frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}y^2} dy$,
> donde μ y σ^2 son la esperanza y varianza común, respectivamente, de $X_1, X_2, \ldots$.

Chebyshev y Lyapunov demostraron estos resultados asumiendo que las variables aleatorias son discretas, es decir, que toman valores únicamente en un conjunto finito o infinito numerable.

Además, durante la segunda mitad del siglo XIX surgió la Mecánica Estadística con los trabajos de Krönig, Clausius, Maxwell y Boltzmann, donde la Teoría de la Probabilidad se constituyó como la herramienta fundamental para el estudio de sistemas con muchas partículas.

También fue en ese periodo cuando surgió la teoría de Mendel sobre la herencia y la teoría de Darwin sobre la evolución de las especies, la primera fundada en un modelo probabilístico y la segunda planteando que el surgimiento de nuevas especies se realiza al azar. Más aún, los estudios de datos crecieron a un ritmo acelerado con los trabajos de Bienaymé, Quetelet y Galton, entre otros.

De esta forma, a finales del siglo XIX el azar y la Teoría de la Probabilidad eran ya parte inseparable del cuerpo científico de la época.

12.5. El Cálculo de Probabilidades durante los primeros 30 años del siglo XX

A pesar del desarrollo que tenía el Cálculo de Probabilidades a finales del siglo XIX, no había una definición satisfactoria de la probabilidad. Eso es lo que afirmaba Henri Poincaré

(1854-1912) en la primera frase del capítulo I de su libro de probabilidad, publicado en 1896 ([**75**]):

> *"No se puede dar una definición satisfactoria de la probabilidad."*

En su libro, enunció la definición clásica de probabilidad:

> *"La probabilidad de un evento es el cociente de los casos favorables a un evento y el número total de casos posibles", aclarando mediante algunos ejemplos que se debe agregar a dicha definición la condición de que todos los casos sean igualmente probables.*

Comentó entonces que:

> *"La definición completa de la probabilidad es una especie de petición de principio: ¿cómo reconocer que todos los casos son igualmente probables? Aquí, una definición matemática no es posible; deberemos, en cada aplicación, hacer convenciones, decir que consideramos tal y tal caso como igualmente probables. Esas convenciones no son completamente arbitrarias, pero escapan al espíritu del matemático que no tendrá más que examinarlas, una vez que son admitidas. Así, todo problema de probabilidad ofrece dos periodos de estudio: el primero, metafísico por así decirlo, el cual legitima tal o cual convención; el segundo, matemático, que aplica a esas convenciones las reglas del cálculo."*

En cuanto al azar, de ser pensado únicamente un producto de nuestra ignorancia pasó a conceptualizarse como algo objetivo. En el mismo libro, Poincaré expresó claramente este cambio:

> *"... en la teoría cinética de los gases, se encuentran las conocidas leyes de Mariotte y de Gay-Lussac, gracias a la hipótesis de que las velocidades de las moléculas gaseosas varían irregularmente, es decir, al azar. Las leyes observables serían mucho menos simples, dirían los físicos, si las velocidades estuvieran arregladas por alguna ley elemental simple, si las moléculas estuvieran, como se dice, organizadas, si obedecieran a alguna disciplina.*

Es gracias al azar, es decir, gracias a nuestra ignorancia, que podemos concluir; y entonces, si la palabra azar es simplemente un sinónimo de ignorancia, ¿qué querría decir eso? ¿Se traduciría entonces como sigue? Me pide usted que le prediga los fenómenos que van a producirse. Si, por desgracia, conociera las leyes de esos fenómenos, podría lograrlo únicamente mediante cálculos inextricables y debería renunciar a responderle; pero, como tengo la suerte de ignorarlas, le voy a responder en seguida. Y, lo más extraordinario, es que mi respuesta será correcta. Se requiere entonces que el azar sea más que el nombre que le damos a nuestra ignorancia."

Agregó Poincaré en su libro básicamente lo que ya había formulado Laplace como las bases del Cálculo de Probabilidades.

Decía Poincaré que el Cálculo de Probabilidades tiene como base dos teoremas: el **teorema de las probabilidades totales** y el **teorema de las probabilidades compuestas**.

$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$.

$P(A \wedge B) = P(B \mid A) P(A)$.

Donde $A \vee B$ representa la ocurrencia de alguno de los dos eventos A y B (incluyendo la ocurrencia de ambos), $A \wedge B$ representa la ocurrencia simultánea de los eventos A y B y $P(B \mid A)$ es la probabilidad de ocurrencia del evento B dado que el evento A ocurre.

En particular, si A y B no pueden ocurrir simultáneamente, entonces:

$P(A \vee B) = P(A) + P(B)$.

De manera más general, si $A_1, A_2, \ldots, A_n$ son eventos tales que ningún par de ellos puede ocurrir simultánemente, entonces:

$P(A_1 \vee A_2 \vee \ldots \vee A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$.

Como lo mencionamos antes, a esta propiedad se le conoce como la **propiedad de la aditividad finita**.

Poincaré recogió en su libro las inquietudes de su época acerca del Cálculo de Probabilidades: No estaba bien fundamentada.

Esta era una inquietud que había no únicamente en relación a la Probabilidad. En el Congreso Internacional de Matemáticas de 1900, David Hilbert (1862-1943) expresó esas inquietudes de la siguiente manera ([**48**]):

> *"Pienso que en cualquier lugar en donde se presenten ideas matemáticas, sea en Filosofía, sea en Geometría, sea en Física, se plantea el problema de la discusión de los principios fundamentales, base de esas ideas, y del establecimiento de un sistema simple y completo de axiomas."*
> *"Las investigaciones sobre los principios fundamentales de la geometría nos conducen a plantear este problema: Tratar con base en ese modelo las ramas de la Física donde las Matemáticas juegan actualmente un papel preponderante; esas ramas de la ciencia son, antes que cualesquiera otras, el Cálculo de Probabilidades y la Mecánica."*

La invención de la Teoría de la Medida a principios del siglo XX vino a resolver el problema de la fundamentación del Cálculo de Probabilidades, surgiendo así un cuerpo teórico, puramente matemático, el cual constituye que ahora podemos llamar la Teoría de la Probabilidad.

Inmediatamente después del surgimiento de la teoría de la medida de Lebesgue, se dio una relación con el Cálculo de Probabilidades.

En 1904 ([**8**]), Émile Borel (1871-1956) planteó que la integral clásica (de Riemann) es insuficiente para tratar algunos problemas de probabilidad :

Si se sabe que un número x está comprendido entre 0 y 1, ¿cuál es la probabilidad de que x sea un número racional?

Utilizando la integral de Riemann, el problema no tiene solución.

Utilizando la integral de Lebesgue, la respuesta es 0.

En un inicio la identificación de la probabilidad con una medida se hizo únicamente en los problemas que caían dentro de un esquema geométrico.

En el año 1909, se publicó un articulo de Borel el cual abrió una polémica acerca de las propiedades que debían pedirse a la función probabilidad en una formulación axiomática.En ese artículo, titulado *Les probabilités dénombrables et leurs applications arithmétiques*, decía Borel ([**9**]):

> *"Se distinguen generalmente, en los problemas de probabilidad, dos categorías principales, dependiendo de que el número de casos posibles sea finito o infinito: la primera categoría constituye lo que se llama las* probabilidades discontinuas*, o probabilidades en el dominio del discontinuo, mientras que la segunda categoría comprende las* probabilidades continuas o probabilidades geométricas*. Tal clasificación aparece como incompleta cuando se consideran los resultados de la Teoría de Conjuntos; entre la potencia de los conjuntos finitos y la potencia del continuo se encuentra la potencia de los conjuntos numerables; me propongo mostrar brevemente el interés respecto a las cuestiones de probabilidad en cuyo enunciado intervienen tales conjuntos; las llamaré, para abreviar,* probabilidades numerables*."*

> Enunciaremos el teorema de Borel utilizando el concepto de **ensayo de Bernoulli**, el cual se define como un experimento aleatorio que admite únicamente dos posibles resultados: éxito y fracaso.

> **Teorema de Borel**. Consideremos una sucesión infinita numerable de ensayos de Bernoulli y sea p_n la probabilidad de éxito en el ensayo n.
> Denotemos por A_∞ al evento:
> A_∞: Se obtiene una infinidad de éxitos.
> Entonces:
> Si la serie $\sum_{n=1}^{\infty} p_n$ es convergente, $P(A_\infty) = 0$.
> Si los ensayos de Bernoulli son independientes y la serie $\sum_{n=1}^{\infty} p_n$ es divergente, $P(A_\infty) = 1$.

En su razonamiento, Borel utilizó algunas de las propiedades que son equivalentes a la σ-aditividad; sin embargo él consideraba que la σ-aditividad no podía considerarse como una propiedad de cualquier función de probabilidad. Para fundamentar su afirmación, daba el siguiente ejemplo:

> *"Supongamos, por ejemplo, que existe una manera de elegir de entre la colección infinita de números enteros, uno de ellos al azar, de manera que cada uno de ellos tenga la misma probabilidad, esta probabilidad deberá entonces ser nula, pero su suma debe ser igual a 1."*

El teorema de Borel tiene ahora una formulación más general, conocida como lema de Borel-Cantelli.

Obsérvese que el teorema de Borel rebasó el marco clásico ya que planteó el cálculo de la probabilidad de eventos cuya ocurrencia o no ocurrencia depende de los resultados de una infininidad de ensayos de Bernoulli.

El artículo de Borel causó un gran impacto en su época sobre todo por una aplicación de sus resultados para deducir una propiedad importante de los números reales.

Sea q es un número natural mayor que 1 y, dado $x \in (0,1)$, expresemos x en la base q:

$$x = \sum_{j=1}^{\infty} \frac{b_j}{q^j},$$

donde cada b_j es un entero no negativo menor que q.

Dado un número $b \in \{0, 1, 2, \ldots, q-1\}$ denotemos por $f_n(b)$ a la fracción que resulta de dividir entre n el número de veces que aparece b en los primeros n términos del desarrollo de x en base q. Cuando $\lim_{n \rightsquigarrow \infty} f_n(b)$ existe, llamemos a ese límite frecuencia total de b en x.

Se dice que $x = \sum_{j=1}^{\infty} \frac{b_j}{q^j}$ es normal con respecto a la base q si dado cualquier número $b \in \{0, \ldots, q-1\}$, la frecuencia total de b en x existe y su valor es igual a $\frac{1}{q}$.

Se dice que $x \in (0,1)$ es absolutamente normal si es normal con respecto a cualquier base $q \in \{2, 3, \ldots\}$.

Borel demostró entonces el siguiente resultado:

Para cada $j \in \mathbb{N}$, seleccionemos al azar un elemento del conjunto $\{0, \ldots, q-1\}$ y definamos x como la serie $\sum_{j=1}^{\infty} \frac{b_j}{q^j}$. Entonces, la probabilidad de que x sea normal con respecto a la base q es igual a 1.

El resultado de Borel puede expresarse en la forma siguiente:

> Sea $\mathcal{E}$ un experimento aleatorio y A un evento relativo a ese experimento, de probabilidad igual a p. Consideremos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento $\mathcal{E}$, de tal manera que cada repetición es independiente de las otras. Sea X_n el número de veces que ocurre el evento A en las primeras n repeticiones del experimento, entonces $P\left[\lim_{n \rightsquigarrow \infty} \frac{X_n}{n} = p\right] = 1$.

La forma general de este resultado se conoce como **Ley Fuerte de los Grandes Números**.

Más tarde, Hausdorff formuló y demostró el resultado de Borel,acerca de los números normales, utilizando La Teoría de la Medida:

Sea q es un número natural mayor que 1, entonces la medida del conjunto de puntos en el intervalo $(0,1)$ que son normales con respecto a la base q, es igual a 1.

Como corolario se tiene que la medida del conjunto de puntos en el intervalo $(0,1)$ que son absolutamente normales es igual a 1.

Sin embargo, hacia el año 1914 todavía no se identificaba a cualquier función de probabilidad con una medida pues ni siquiera estaba desarrollada la teoría general de la medida en espacios abstractos. En ese momento se contaba ya con la teoría de integración de Lebesgue y la correspondiente teoría de la medida en $\mathbb{R}^n$ y eran entonces éstas las únicas medidas que al normalizarlas se consideraban probabilidades. Esto es lo que hizo Felix Hausdorff en su libro, publicado en 1914 ([**?**]). Ahí consideró que si A y B son dos conjuntos medibles de medida finita y $A \subset B$, entonces la medida de A dividida entre la medida de B puede considerarse como la probabilidad de que un punto que se selecciona en el conjunto B pertenezca al conjunto A. También en ese libro Hausdorff demostró el teorema de Borel sobre los números normales dentro del marco de la teoría de la medida.

En el libro de Hausdorff de 1914 se considera a la probabilidad como un ejemplo y una aplicación de la teoría de la medida. Hausdorff no identificaba a una probabilidad con una medida, pero mostró que una medida normalizada tiene todas las propiedades de una probabilidad.

El libro de Hausdorff fue durante mucho tiempo la referencia estándar para la teoría de conjuntos; entonces la conexión entre la probabilidad y la teoría de la medida puede considerarse como bien establecida en la literatura matemática desde 1914.

Por otra parte, en 1913, Johann Radon había ya desarrollado una teoría general de la medida en $\mathbb{R}^n$ ([**76**]) y en 1915, con base en el trabajo de Radon, Maurice René Fréchet extendió la teoría de la medida a espacios abstractos, definiendo las funcionales aditivas ([**36**]). De esta manera, se puede decir que, en ese momento, aunque posteriormente todavía se demostrarían algunos resultados importantes, ya se contaba con lo básico de una teoría general de la medida.

Sin embargo, hacia 1915, aunque Frechet ya había desarrollado una teoría de la medida en espacios abstractos, no podía hacerse una identificación automática de una función de probabilidad con una medida mientras no se resolviera el problema de la existencia de una medida asociada a cada problema de probabilidad.

En 1914, Carathéodory ([**16**]) dio un método para construir medidas en $\mathbb{R}^n$ vía una medida exterior y este método puede extenderse al caso de medidas en espacios abstractos. Sin embargo, la definición de medidas en espacios de dimensión infinita no es un problema que se haya resuelto inmediatamente después del trabajo de Fréchet sobre la definición general de una medida.

Fue P.J. Daniell quien entre 1918 y 1920 desarrolló una teoría de integración en espacios de dimensión infinita ([**21**], [**22**], [**23**], [**24**]). Daniell no se basó para esto en el resultado de Carathéodory sino que desarrolló su propio método.

Básicamente el método de Carathéodory para definir una medida consiste en partir de una medida definida sobre un álgebra de subconjuntos de un conjunto dado Ω y en extender esta medida a una σ-álgebra que contiene a los conjuntos del álgebra de la que se partió. En cambio, el método de Daniel consiste en partir de una integral. definida para una cierta familia de funciones y en extender esta integral a una familia suficientemente grande de funciones. Los dos métodos son equivalentes en el sentido de que una vez teniendo una medida se puede definir una integral e inversamente, una vez teniendo una integral se puede definir una medida.

Por otra parte, el estudio de los teoremas límite había puesto en el centro de la atención de los probabilistas a las variables aleatorias. El estudio de las variables aleatorias condujo a Richard Edler Von Mises (1883-1953) a identificar, en el año 1919, una ley de probabilidad con la función de distribución ([**94**], [**95**]). Esta misma identificación la hizo Paul Pierre Lévy (1886-1971) en su libro *Calcul des Probabilités*, publicado en 1925 ([**62**]), donde, además, identificaba a una función de distribución con una medida sobre $\mathbb{R}$ y a una función de distribución conjunta con una medida sobre $\mathbb{R}^n$. De esta forma, dada una sola variable aleatoria, se puede asociar a ésta una medida sobre $\mathbb{R}$; dado un número finito de variables aleatorias, se puede asociar a esa familia una medida sobre $\mathbb{R}^n$, para alguna n.

Pero, ¿**cómo asociarle una medida a una familia infinita de variables aleatorias**?

Algunos resultados parciales consistentes en asociar una medida a una familia infinita de variables aleatorias se encuentran en los trabajos de Hugo Dyonizy Steinhaus (1887-1972) ([**86**]) y de Norbert Wiener (1894-1964) ([**96**], [**97**], [**98**], [**99**], [**100**], [**101**], [**102**]).

En 1923, Steinhaus consideró una sucesión infinita de ensayos de Bernoulli, en cada uno de los cuales la probabilidad de éxito es $\frac{1}{2}$, y las variables aleatorias, $X_1, X_2, \ldots$, son tales que:

$$X_j = \begin{cases} 1 & \text{si hay éxito en el ensayo } j \\ 0 & \text{si no lo hay} \end{cases}$$

El conjunto de posibles resultados del experimento aleatorio así definido consiste entonces del conjunto de sucesiones de 0's y 1's, el cual se puede poner en correspondencia, excepto por un conjunto numerable, con el intervalo $[0, 1]$.

Definió la axiomática para el juego de cara o cruz dándole a la función de probabilidad la propiedad de σ-aditividad.

Mostró entonces que comenzando por asignar probabilidades a eventos que dependen únicamente de un número finito de ensayos, las propiedades que dio a la función de probabilidad permiten definirla (extenderla) para todos los subconjuntos Lebesque-medibles y que la medida que se obtiene es precisamente la medida de Lebesgue.

Steinhaus consideró también el problema de la convergencia de series aleatorias de la forma $\sum_{n=1}^{\infty} \pm c_n$, en donde cada c_n es un número real y el signo de c_n se elige al azar.

Su modelo nuevamente consiste en identificar una sucesión infinita de signos como un punto del intervalo $[0, 1]$ y entonces nuevamente asumiendo que la función de probabilidad es $\sigma-$aditiva, mostró que la función de probabilidad es la medida de Lebesgue sobre los conjuntos Lebesgue-medibles. Con base en esto demostró que la probabilidad de convergencia de una serie así definida necesariamente es 0 ó 1.

En 1924, Norbert. Wiener consideró también el problema de la convergencia de series aleatorias, pero su método fue distinto al de Steinhaus.

Wiener trabajaba con funcionales lineales sobre espacios de funciones y seguía el método de Daniell para extender tales funcionales:

Sea Ω es el conjunto de todas las sucesiones posibles de signos. Si φ es una función definida sobre Ω cuyos valores dependen únicamente de los primeros n signos para alguna n, Wiener definió $I(\varphi)$ como el promedio de los 2^n valores que toma φ dependiendo de los primeros n signos de la sucesión. Demostró entonces que esa funcional así definida satisface las propiedades del teorema de extensión de Daniell, de manera que dicha funcional se puede extender de manera única al conjunto de todas las funciones medibles.

Con el mismo método, entre 1921 y 1923, construyó un modelo matemático para el movimiento browniano, para lo cual definió una medida de probabilidad $\sigma-$aditiva sobre el espacio de las funciones continuas. Es este trabajo el que marcó la pauta para poder definir una medida asociada a cualquier problema de probabilidad.

El **movimiento browniano** consiste en el movimiento de un grano de polen que se coloca sobre agua. En el año 1827, al estudiar el proceso de fertilización de las flores de varias plantas, Robert Brown observó que los granos de polen se movían.

Lo que hizo Wiener fue construir una medida de probabilidad sobre el conjunto de las posibles trayectorias que puede seguir una de esas partículas colocadas sobre un fluido.

Así que, pesar de la objeción de Borel, se volvió cada vez más frecuente asumir como válida ya sea la propiedad de σ-aditividad de la función de probabilidad o bien alguna de sus formas equivalentes.

Para el año 1925 algunos autores aceptaban ya a la σ-aditividad como una propiedad general de la función de probabilidad y entonces consideraban a la probabilidad como una medida. Esto queda claro en el libro de Paul Pierre Lévy de 1925, donde, además, se define a la probabilidad en forma axiomática.

Un año antes se publicó un artículo de Lévy titulado *Les lois de probabilité dans les ensembles abstraits*, en el cual dice ([**61**]):

> *Una ley de probabilidad será naturalmente bien definida en un conjunto abstracto E si se conoce la probabilidad de todo subconjunto de E. Esta probabilidad deberá gozar de las propiedades siguientes:*
>
> (i) *A dos conjuntos V_1 y V_2 sin elementos comunes y al conjunto V constituido por su unión, corresponden números α_1, α_2 y α tales que $\alpha = \alpha_1 + \alpha_2$.*
>
> (ii) *Un enunciado análogo es verdadero si se considera una infinidad numerable de conjuntos $V_1, V_2, \ldots$, sin puntos comunes dos a dos.*
>
> (iii) *Los valores de α son siempre positivos o nulos y al conjunto E completo corresponde un valor igual a la unidad.*

> Decía Lévy que, utilizando el lenguaje del Cálculo Funcional, α es una funcional aditiva en el sentido de Fréchet (es decir, una medida).

Agregaba después que en la práctica se considera una ley de probabilidad como definida sin que la probabilidad α esté definida para todos los subconjuntos de E. Cita para esto el caso en que la probabilidad de un subconjunto del intervalo $[0, 1]$ está dada por su medida de Lebesgue, en cuyo caso la probabilidad únicamente está definida para los conjuntos medibles.

Como puede verse, Lévy formuló aquí la Teoría de la Probabilidad en su forma axiomática moderna. Sin embargo, aunque en ese artículo Lévy formuló un método para construir medidas en espacios de dimensión infinita, éste no era lo suficientemente general.

Sin lugar a dudas, Lévy fue el más grande probabilista del siglo XX. Publicó más de 100 artículos acerca del Cálculo de Probabilidades y los Procesos Estocásticos. Su trabajo lo sistematizó en 3 libros: 1. El ya mencionado libro de 1925, *Calcul des Probabilités*, donde, además de lo que ya dijimos antes, realizó una gran sistematización del Cálculo de Probabilidades e hizo ver toda la fuerza que tiene la función característica para tratar los teoremas límite. 2. *Théorie de l'addition des variables aleatoires*, publicado en 1937, el cual contiene una amplia discusión sobre el concepto de probabilidad y el primer estudio sistemático sobre las distribuciones infinitamente divisibles. 3. *Processus Stochastiques et Mouvement Brownien*, publicado en 1948, donde introdujo los procesos con incrementos independientes e hizo un estudio minucioso del movimiento browniano, probando resultados que sólo pudieron ser demostrados formalmente años más tarde, utilizando el Cálculo Estocástico. Se mantuvo publicando artículos acerca de estos temas hasta 1970, un año antes de su fallecimiento, a los 85 años de edad.

Paralelamente a lo anterior, el Cálculo de Probabilidades tuvo una gran difusión tanto por parte de la escuela francesa como de la escuela rusa. En particular, bajo la dirección de Émile Borel, se publicó una colección de libros acerca del Cálculo de Probabilidades y sus aplicaciones, la cual consta de 4 tomos: I. *Les principes de la théorie des probabilités.* II. *Les applications de la théorie des probabilités aux sciences mathématiques et aux sciences physiques.* III. *Les applications de la théorie des probabilités aux sciences économiques et aux sciences biologiques.* IV. *Applications diverses et conclusion.* Esta colección está compuesta por 19 libros, los cuales fueron publicados entre 1924 y 1939:

Tome III, Fascicule 1. Assurances sur la vie. Calcul des primes, par Henri Galbrun (1924).

Tome I, Fascicule 1. Principes et formules classiques du calcul des probabilités, par Émile Borel et rédigé par René Lagrange (1925).

Tome II, Fascicule 3. Mécanique statistique classique, par Émile Borel et rédigé par Francis Perrin, (1925).

Tome II, Fascicule 1. Applications à l'arithmétique et à la théorie des fonctions, par Émile Borel et rédigé par Paul Dubreil (1926).

Tome II, Fascicule 2. Probabilités géométriques, par Robert Deltheil (1926).

Tome IV, Fascicule 1. Applications au tir, par Jules Haag (1926).

Tome III, Fascicule 2. Assurances sur la vie. Calcul des réserves, par Henri Galbrun (1927).

Tome I, Fascicule 2. Erreurs et moindres carrés, par René Deltheil (1930).

Tome II, Fascicule 4. Applications de la théorie des probabilités à l'astronomie, par Carl V. L. Charlier (1931).

Tome III, Fascicule 3. Applications de la statistique à la démographie et à la biologie, par René Risser (1932).

Tome I, Fascicule 4. Les principes de la statistique mathématique, par René Risser et Claude-Émile Traynard (1933).

Tome III, Fascicule 4. Théorie mathématique de l'assurance invalidité et de l'assurance nuptialité. Définitions et relations fondamentale, par Henri Galbrun (1933).

Tome III, Fascicule 5. Théorie mathématique de l 'assurance invalidité et de l'assurance nuptialités. Calcul des primes et des réserves, par Henri Galbrun (1933).

Tome III, Fascicule 6. Théorie mathématique de l'assurance maladie, par Henri Galbrun (1934).

Tome I, Fascicule 3. Recherches théoriques modernes sur la théorie des probabilités, par Maurice Fréchet. Premier Livre: Généralités sur les Probabilités. Variables aléatoires (avec une note de Paul Lévy) (1937).

Tome I, Fascicule 3. Recherches théoriques modernes sur la théorie des probabilités, par Maurice Fréchet. Deuxième Livre : Méthode des fonctions arbitraires, théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles (1938).

Tome IV, Fascicule 2. Applications au jeux de hasard, par Émile Borel et rédigé par Jean Ville (1938).

Tome II, Fascicule 5. Mécanique statistique quantique, par Francis Perrin, (1939).

Tome IV, Fascicule 3. Valeur pratique et philosophie des probabilités, par Émile Borel (1939).

12.6. La axiomática

Como lo mencionamos anteriormente, Lévy, en su libro de 1925, asumía como válida la σ-aditividad para cualquier función de probabilidad. Algunos años después, se utilizaba ya para demostrar formas generales de los teoremas límite.

En el año 1930, para probar la ley fuerte de los grandes números, Andrey Nikolaevich Kolmogorov (1903-1987) ([**52**]) utilizó la propiedad de σ-subaditividad de la función de probabilidad, la cual es equivalente a la σ-aditividad. Además, Kolmogorov utilizó el hecho de que la unión numerable de eventos de probabilidad cero tiene también probabilidad cero, la cual también es consecuencia de la σ-subaditividad.

> Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ. Entonces:
> $P\left[\lim_{n\rightsquigarrow\infty} \frac{X_1+\cdots+X_n}{n} = \mu\right] = 1.$

Sin embargo, la polémica sobre la propiedad de $\sigma-$aditividad de la función de probabilidad continuaba. Resalta en esta polémica una serie de artículos que publicaron Maurice Fréchet y Bruno de Finetti en el año 1930 ([**25**], [**26**], [**27**], [**40**], [**41**]).

De Finetti consideraba que se llega a contradicciones cuando se admite la extensión del teorema sobre las probabilidades totales al caso de una sucesión infinita de eventos mutuamente excluyentes. Como ejemplo consideraba una variable aleatoria X la cual únicamente puede tomar valores en el conjunto infinito $\{\varepsilon_1, \varepsilon_2, \ldots\}$ de tal forma que todos ellos son igualmente probables. Los eventos $[X = \varepsilon_i]$ tienen entonces probabilidad cero, pero su unión tiene probabilidad 1.

Fréchet argumentaba que él ya había señalado, en sus cursos y en una memoria que se encontraba en prensa, que efectivamente la extensión del teorema sobre las probabilidades totales al caso de una sucesión infinita de eventos no es una consecuencia inevitable de los principios generales admitidos en las bases del Cálculo de Probabilidades. Pero agregaba que de Finetti únicamente había visto una de las dos alternativas: "si sus ejemplos tienen sentido, entonces tal extensión no es posible. pero la otra alternativa es que si tal extensión es posible entonces los ejemplos no tienen sentido." Fréchet prefería entonces asumir que los ejemplos de de Fineti no tienen sentido, en particular consideraba, con relación al mencionado ejemplo de de Fineti, que es imposible suponer que los posibles valores de X son igualmente probables. Continuaba argumentando que la misma alternativa se presenta en la teoría de la medida de Lebesgue, donde se tiene que restringir la familia de conjuntos a los cuales se les puede asignar una medida pues no todos los conjuntos resultan ser medibles. De la misma manera, en el ejemplo de de Fineti no es posible asignarle una probabilidad a los conjuntos $[X = \varepsilon_i]$ de tal manera que todas ellas sean iguales.

De Fineti respondió con nuevas objeciones. Se preguntaba si los eventos que se tienen que excluir de aquellos a los cuales se asigna una probabilidad no son tan interesantes como éstos últimos. Para él Fréchet únicamente evitaba formalmente la dificultad y se seguía preguntando: ¿Es admisible excluir la concepción de una infinidad de eventos mutuamente excluyentes que sean igualmente probables?

Fréchet contraargumentó que las contradicciones a que hace referencia de Fineti son familiares para todos aquellos al corriente en la teoría de la medida. En cuanto al interés que pueden tener los conjuntos no medibles responde que en realidad no se presentan en las aplicaciones. En cuanto a la necesidad de excluir algunas medidas como posibles, consideraba, por ejemplo, que se puede pensar en asignar una medida igual a 1 a toda la recta real, una medida igual a $\frac{1}{2}$ a toda semirecta, una medida igual a $\frac{1}{3}$ a todos los conjuntos formados por la unión de una sucesión infinita de intervalos de longitud λ, de tal manera que cada par de ellos esté separado por un intervalo de longitud 2λ, etc. En ese caso, toda la recta real sería la unión de una sucesión de intervalos consecutivos cuyas medidas tendrían que ser nulas, de manera que su suma no podría ser igual a 1. Por lo tanto, se debe de excluir la concepción de medidas iguales de esos intervalos o bien se deben de considerar como no medibles.

Resulta aquí claro que para Fréchet la probabilidad era siempre una medida, aún a costa de tener que excluir algunos experimentos aleatorios que pueden ser definidos formalmente, aunque también resultaba claro para él que ésta es únicamente una alternativa que se puede elegir, pero que no era aceptada por todos en ese momento. Esta posición resulta todavía más evidente al argumentar en contra de otra objeción que hace de Fineti en su segundo artículo. Decía de Fineti que no se debe eludir una dificultad de principio mediante una convención y que una vez puesta la definición de probabilidad, de una manera conforme a nuestra intuición, si esta definición permite atribuir un valor a la probabilidad de uno de los eventos clasificados como no probabilizables, no se tiene el derecho de excluir ese evento. Fréchet respondió entonces que la principal dificultad en el argumento de de Fineti reside en el hecho de que, hasta ese momento, ninguna definición de la probabilidad había obtenido una adhesión general. Agregaba que si se adopta el punto de vista axiomático, la solución es inmediata y consiste en poner como postulado el principio de las probabilidades totales en su forma completa (es decir, la propiedad de aditividad numerable).

Finalmente, en el año 1933, Kolmogorov publicó un artículo titulado *Foundations of the Theory of Probability* ([**53**]) en el cual estableció la formulación de la Teoría de la Probabilidad que prevalece hasta nuestros días. Dice ahí:

> *"Después de las publicaciones de las investigaciones de Lebesgue, las analogías entre medida de un conjunto y probabilidad de un evento y entre la integral de una función y la esperanza matemática de una variable aleatoria se hicieron evidentes. Pero para que la teoría de la probabilidad pudiera basarse en tales analogías era todavía necesario hacer las teorías de la medida y de la integración independientes de los elementos geométricos los cuales estaban en el trasfondo con Lebesgue. Esto ha sido hecho por Fréchet. Mientras que una concepción de la teoría de la probabilidad basada sobre el punto de vista general citado antes se ha dado durante algún tiempo entre ciertos matemáticos, estaba faltando una exposición completa de todo el sistema, libre de extrañas complicaciones."*

Estableció entonces como modelo matemático de un fenómeno probabilístico una terna $(\Omega, \Im, P)$, donde Ω es un conjunto, $\Im$ una $\sigma-$álgebra de subconjuntos Ω y P una medida de probabilidad definida sobre $\Im$.

Con este modelo Kolmogorov logró entonces articular los diferentes conceptos de la teoría de la probabilidad, como el de probabilidad condicional y la independencia de eventos y de variables aleatorias. Mostró además como los resultados fundamentales de la teoría de la probabilidad se articulan en un enfoque axiomático, exponiendo, dentro de este nuevo contexto, las leyes débil y fuerte de los grandes números.

En su monografía, Kolmogorov introdujo el concepto de esperanza condicional, con lo cual mostró como el enfoque axiomático basado en la Teoría de la Medida aporta a la Teoría de la Probabilidad poderosas herramientas.

Finalmente, Kolmogorov, utilizando el método de Carathéodory, dio un método general, además de simple, para construir medidas de probabilidad en espacios de dimensión infinita:

> Dada cualquier familia de variables aleatorias, partiendo de sus distribuciones finito dimensionales, es posible construir un espacio de probabilidad $(\Omega, \Im, P)$ de tal manera que la medida P restringida a los eventos que dependen únicamente de un número finito de las variables aleatorias dadas coincide con la determinada por la distribución finito-dimensional correspondiente.

Después del trabajo de Kolmogorov la aceptación de probabilidad como una medida fue unánime.

12.7. Acerca de la propiedad de σ-aditividad de la función de probabilidad

El considerar a la probabilidad como una medida (σ-aditiva) constituye únicamente una elección; bien podría elegirse definir un espacio de probabilidad como una terna $(\Omega, \Im, P)$, donde $\Im$ es un álgebra de subconjuntos de Ω y P una función finitamente aditiva definida sobre $\Im$. Tendríamos así un modelo matemático similar al de la terna de Kolmogorov y la teoría se podría desarrollar con base en ese modelo. En cualquiera de los dos casos, con una función de probabilidad σ-aditiva o con una únicamente finitamente aditiva, la teoría que se desarrolle es puramente matemática y los resultados que se obtengan son válidos, matemáticamente, dentro de esa teoría. En los dos casos se trata únicamente de un modelo matemático de los fenómenos aleatorios. El fenómeno aleatorio en sí mismo no no es matemático, ni contiene matemática alguna. En Cambio, el modelo matemático es una abstracción producto del pensamiento humano; es parte del simbolismo que el ser humano ha creado para tratar de entender los fenómenos naturales.

El preguntarse si una función de probabilidad es o no σ-aditiva no tiene sentido, o si se quiere, es una trivialidad. Si en el modelo que estamos utilizando tomamos a la probabilidad como σ-aditiva, entonces lo es; de otra forma, no lo es.

Lo que hizo Kolmogorov no fue demostrar que toda función de probabilidad es σ-aditiva. Recapitulemos el proceso:

Primero, a una variable aleatoria se le asocia (de hecho se le identifica) con una función (llamada función de distribución) no decreciente, la cual resulta ser continua por la derecha, gracias a que se asume que la función de probabilidad es σ-aditiva; dicho de otra forma, la σ-aditividad es parte de los axiomas, de manera se asume a priori como válida. De ahí, como decíamos, la función de distribución resulta ser continua por la derecha. De ahí que podamos definir una medida (σ-adtiva) a partir de esa función de distribución y, de hecho, se identifica a la función de distribución con la medida que genera. De esta forma, asociada con una variable aleatoria real X se construye una medida sobre los subconjuntos borelianos de $\mathbb{R}$, la cual representa a la variable aleatoria. Pero, reiteramos, esto puede hacerse gracias a que de inicio se asume como válida la σ-aditividad de la función de probabilidad.

Después se trata el caso de un número finito de variables aleatorias, X_1, X_2, $\ldots$, X_n, a las cuales se les asocia una función de distribución conjunta, a partir de la cual se puede definir una medida sobre los subconjuntos borelianos de $\mathbb{R}^n$. Nuevamente esto es posible ya que se asume que la función de probabilidad es σ-aditiva.

El siguiente paso consiste en considerar una infinidad (numerable o no numerable) de variables aleatorias y Kolmogorov demostró que, partiendo de que cada subconjunto finito, de esa infinidad de variables aleatorias, tiene asociada una función de distribución y, por lo tanto, una medida sobre los subconjuntos borelianos de $\mathbb{R}^n$, para alguna $n \in \mathbb{N}$, es posible definir una medida sobre algún espacio medible, la cual representa a la infinidad de variables aleatorias.

Lo que dice entonces el teorema de Kolmogorov es que la propiedad de σ-aditividad es consistente en el sentido de que si se asume como valida para el caso finito, entonces la σ-aditividad se puede extender al caso infinito.

De hecho, el problema de tomar o no la σ-aditividad para la función de probabilidad es el mismo que el que se plantea en la teoría de la medida desarrollada por Lebesgue

Recordemos que Lebesgue planteó el problema de encontrar una función m definida sobre todos los subconjuntos acotados de números reales y satisfaciendo las siguientes condiciones:

(i) m es no negativa.
(ii) m es σ-aditiva.
(iii) $m([0,1]) = 1$
(iv) m es invariante bajo traslaciones.

Lebesgue logró definir una función m, única, definida sobre una σ-álgebra de subconjuntos de $\mathbb{R}$, la cual asigna a cada intervalo su longitud; sin embargo, Vitali demostró que esa σ-algebra no está formada por todos los subconjuntos de los números reales.

Stefan Banach y Kazimierz Kuratowski se plantearon en 1929 el problema de encontrar una función m definida sobre todos los subconjuntos el intervalo $[0,1]$ y satisfaciendo las siguientes condiciones:

(i) m es no negativa.
(ii) m es σ-aditiva.
(iii) Si I es un intervalo, entonces $m(I)$ es igual a la longitud de I.

El resultado de Banach y Kuratowski fue que tal problema no tiene solución ([**?**]). Esto sorprendió a muchos, por ejemplo a Paul Pierre Lévy quien pensaba que, al quitar a la medida la condición de ser invariante bajo traslaciones, es posible asignar una medida a todos los subconjuntos de números reales ([**?**]).

Alfred Tarski atacó en 1930 el problema de la medida en sentido amplio planteándose el problema de encontrar una función m definida sobre todos los subconjuntos del intervalo $[0,1]$ y satisfaciendo las siguientes condiciones:

(i) m es no negativa.
(ii) m es finitamente aditiva.
(iii) Si I es un intervalo, entonces $m(I)$ es igual a la longitud de I.

Tarski mostró que tal problema, así como el análogo en dos o más dimensiones, sí tiene solución. Sin embargo, mostró también que la solución no es única ([**?**]).

El resultado de Tarski se puede extender al caso de una función finitamente aditiva definida sobre un álgebra:

Sean $\mathcal{A}$ y $\mathcal{B}$ dos álgebras de subconjuntos de un conjunto Ω tales que $\mathcal{A} \subset \mathcal{B}$ propiamente, y $P : \mathcal{A} \mapsto [0,1]$ una función finitamente aditiva tal que $P(\Omega) = 1$. Entonces existe una función finitamente aditiva $\overline{P} : \mathcal{B} \mapsto [0.1]$, la cual es una extensión de P. Tal extensión no necesariamente es única. ([**?**])

Por otra parte, el teorema de Caratheodory permite extender, una quasi medida de probabilidad definida sobre un álgebra de subconjuntos de un conjunto Ω a una medida definida sobre la σ-álgebra generada por el álgebra, y tal extensión es única.

Recordemos que cuando se tiene un fenómeno aleatorio, el proceso para asignar probabilidades a los subconjuntos del espacio muestral consiste en partir de la asignación de probabilidades a una determinada familia de subconjuntos y después en extender la función de probabilidad a una familia de subconjuntos del espacio muestral tan grande como sea posible. La σ-aditividad nos permite realizar esa extensión, de manera única, a una familia suficientemente grande de subconjuntos del espacio muestral.

Con base en lo anterior, podemos decir que elegir la σ-aditividad como propiedad de cualquier función de probabilidad nos permite definir, de manera única, la probabilidad de cada evento, a costa de que la familia de eventos tal vez no esté formada por todos los subconjuntos del espacio muestral Ω, mientras que eligiendo únicamente la aditividad finita, podemos definir la probabilidad de cualquier subconjunto del espacio muestral, pero de diferentes maneras.

Cabe decir que si $\Omega = \{\omega_1, \omega_2, \ldots\}$ es un conjunto infinito numerable, $\Im$ el conjunto potencia de Ω y $P : \Im \mapsto [0,1]$ es una función finitamente aditiva tal que $\sum_{k=1}^{\infty} P(\{\omega_k\}) = 1$, entonces P necesariamente es σ-aditiva. En efecto; sea $A = \{\omega_{j_1}, \omega_{j_2}, \ldots\}$ cualquier subconjunto de Ω infinito numerable y $A^c = \{\omega_{i_1}, \omega_{i_2}, \ldots\}$. En vista de que P es no negativa y finitamente aditiva, es monótona no decreciente, así que $P(A) \geq \sum_{k=1}^{n} P(\{\omega_{j_k}\})$ para toda $n \in \mathbb{N}$. Por lo tanto, $P(A) \geq \sum_{k=1}^{\infty} P(\{\omega_{j_k}\})$. De la misma manera, se obtiene $P(A^c) \geq \sum_k P(\{\omega_{i_k}\})$. Además, como la serie $\sum_{k=1}^{\infty} P(\{\omega_k\})$ converge a 1, las dos series $\sum_{k=1}^{\infty} P(\{\omega_{j_k}\})$ y $\sum_k P(\{\omega_{i_k}\})$ son también convergentes y su suma es igual a 1. Ahora bien, si $P(A) > \sum_{k=1}^{\infty} P(\{\omega_{j_k}\})$, entonces tendríamos:

$$1 \geq P(\Omega) = P(A) + P(A^c) > \sum_{k=1}^{\infty} P(\{\omega_{j_k}\}) + \sum_k P(\{\omega_{i_k}\}) = 1$$

lo cual es una contradicción. Por lo tanto, podemos concluir que $P(A) = \sum_{k=1}^{\infty} P(\{\omega_{j_k}\})$.

Sea ahora $A_1, A_2, \ldots$ una colección infinita numerable de subconjuntos no vacíos de Ω tales que $A_i \cap A_j = \emptyset$ para $i \neq j$. Si $A = \cup_{n=1}^{\infty} A_n = \{\omega_{k_1}, \omega_{k_2}, \ldots\}$, entonces, como P es no negativa y finitamente aditiva, se tiene $P(A) \geq \sum_{n=1}^{n_0} P(A_n)$ para toda $n_0 \in \mathbb{N}$. Así que:

$$P(A) \geq \sum_{n=1}^{\infty} P(A_n)$$

Por otro lado, dada $N \in \mathbb{N}$, existe $n_0 \in \mathbb{N}$ tal que $\{\omega_{k_j} : j \leq N\} \subset \cup_{n=1}^{n_0} A_n$, así que:

$$\sum_{j=1}^{N} P\left(\left\{\omega_{k_j}\right\}\right) = P(\{\omega_{k_j} : j \leq N\}) \leq P(\bigcup_{n=1}^{n_0} A_n) = \sum_{n=1}^{n_0} P(A_n) \leq \sum_{n=1}^{\infty} P(A_n)$$

Es decir, $\sum_{j=1}^{N} P\left(\left\{\omega_{k_j}\right\}\right) \leq \sum_{n=1}^{\infty} P(A_n)$ para toda $N \in \mathbb{N}$. Por lo tanto:

$$P(A) = \sum_{j=1}^{\infty} P\left(\left\{\omega_{k_j}\right\}\right) \leq \sum_{n=1}^{\infty} P(A_n)$$

Así que, $P(A) = \sum_{n=1}^{\infty} P(A_n)$, lo cual prueba la σ-aditividad.

Por otra parte, debe de observarse que, en general, la σ-aditividad no es una consecuencia de la aditividad finita. Consideremos, por ejemplo, el álgebra $\mathcal{A}$ formada por los subconjuntos de los números naturales que son finitos o de complemento finito y definamos la función $P : \mathcal{A} \mapsto [0,1]$ por $P(A) = 0$ si A es finito y $P(A) = 1$ si A^c es finito. Tal función es finitamente aditiva pero no σ-aditiva. Más aún, por el resultado 12.7, enunciado con anterioridad, P puede extenderse (no de manera única) a una función finitamente aditiva definida sobre la familia de todos los subconjuntos de los números naturales. Tal extensión, la cual está definida sobre una σ-álgebra, resulta entonces ser finitamente aditiva, pero no σ-aditiva.

CAPÍTULO 13

FORMULACIÓN AXIOMÁTICA DE LA TEORÍA DE LA PROBABILIDAD

13.1. Espacios de Probabilidad

Como ya lo mencionamos, la Teoría de la Probabilidad se utiliza para modelar y estudiar fenómenos aleatorios, los cuales tienen la característica de evolucionar de una manera azaroza o que, aún no siendo azarozo su desarrollo, su estudio puede realizarse pensándolo como si lo fuera. Sin embargo, una vez que se formula la Teoría de la Probabilidad en forma axiomática, los elementos que la componen no requieren de una interpretación práctica. La teoría matemática puede desarrollarse a partir de los axiomas, sin hacer referencia a algún fenómeno aleatorio. Se investiga acerca de las propiedades del modelo matemático y se van introduciendo definiciones de nuevos conceptos, dentro del modelo, enriqueciendo así el modelo mismo y demostrando nuevas propiedades. Nos referiremos a este proceso como el **desarrollo formal de la teoría**. Los nuevos conceptos que se van introduciendo provienen en general de problemas que se plantean en el estudio de algún fenómeno natural y las propiedades que se van encontrando del modelo se utilizan para estudiar el fenómeno en consideración; sin embargo el desarrollo formal de la teoría se va dando sin aludir a los problemas que motivan las definiciones de nuevos conceptos.

En lo que sigue vamos a exponer el desarrollo formal de la Teoría de la Probabilidad, así que comenzaremos con las definiciones de los conceptos básicos.

DEFINICIÓN 13.1. *Llamaremos espacio de probabilidad a una terna* $(\Omega, \Im, P)$*, donde* Ω *es un conjunto,* $\Im$ *una* σ*-álgebra de subconjuntos de* Ω *y* P *una medida sobre* $\Im$ *tal que* $P(\Omega) = 1$*, a la cual llamaremos medida de probabilidad. A* Ω *lo llamaremos el espacio muestral, a los elementos de* $\Im$ *eventos y a la medida* P *de un evento* A *la probabilidad de* A*.*

Considerando que cualquier espacio de medida se puede completar, asumiremos que cualquier medida de probabilidad con la que trabajemos es completa.

En el resto de este capítulo asumiremos que tenemos definido un espacio de probabilidad $(\Omega, \Im, P)$.

Recordemos que, antes de su formulación axiomática, el Cáculo de Probabilidades tenía como base dos reglas, la de la probabilidad total y la de las probabilidades compuestas. La primera de ellas queda comprendida en las propiedades de la medida de probabilidad. La segunda, en cambio, no queda contemplada debido a que había un problema en su formulación: se hablaba de probabilidades condicionales sin haber definido lo que eso significaba matemáticamente, de manera que se utilizaba únicamente cuando las probabilidades condicionales con las que se trataba tenían un sentido intuitivo. Lo que se hizo entonces fue formular una definición, en lugar de un teorema o una regla.

Definición 13.2 (**Probabilidad condicional**). *Sean A y B dos eventos y supongamos $P(A) > 0$, se define la probabilidad condicional de B, dada la ocurrencia de A, $P(B|A)$, mediante la fórmula:*

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

La probabilidad condicional dado un evento A es una nueva medida de probabilidad, la cual asigna el valor 1 a A y el valor 0 a A^c; es decir, se trata de una medida de probabilidad concentrada en A. Se puede decir que al tomar probabilidades condicionales dado un evento A, el espacio muestral se reduce, convirtiéndose A en un nuevo espacio muestral.

Definición 13.3 (**Independencia de eventos**). *Diremos que los eventos de una familia no vacía cualquiera $\{A_\gamma\}$, finita o infinita, son estocásticamente independientes si dada cualquier subcolección finita de ellos, $A_{\gamma_1}, \ldots, A_{\gamma_n}$, donde $n \in \mathbb{N}$, se tiene:*

$$P(A_{\gamma_1} \cap \cdots \cap A_{\gamma_m}) = P(A_{\gamma_1}) \cdots P(A_{\gamma_m}).$$

Definición 13.4 (**Eventos mutuamente excluyentes**). *Diremos que los eventos de una familia no vacía cualquiera $\{A_\gamma\}$, finita o infinita, son mutuamente excluyentes si cualquier par de ellos son conjuntos ajenos.*

Ahora, algunas propiedades simples:

Proposición 13.1. *Si los eventos de una familia no vacía $\{A_\gamma\}_{\gamma\in\Gamma}$ son mutuamente excluyentes y Γ' es cualquier subconjunto no vacío de Γ, entonces los eventos de la familia $\{A_\gamma\}_{\gamma\in\Gamma'}$ son mutuamente excluyentes.*

Proposición 13.2. *Si los eventos de una familia no vacía $\{A_\gamma\}_{\gamma\in\Gamma}$ son independientes y Γ' es cualquier subconjunto no vacío de Γ, entonces los eventos de la familia $\{A_\gamma\}_{\gamma\in\Gamma'}$ son independientes.*

Proposición 13.3. *Sea $\{A_1, , \ldots, A_n\}$ una familia de eventos independientes, donde $n \geq 2$, y reemplacemos uno de ellos, cualquiera, por su complemento, entonces los eventos de la nueva familia siguen siendo independientes.*

Demostración

Reordenemos los eventos $A_1, , \ldots, A_n$ de tal manera que el evento que reemplazamos por su complemento sea A_n.

Sea $\{n_1, \ldots, n_k\}$ cualquier subconjunto no vacío del conjunto $\{1, \ldots, n\}$, donde los elementos $n_1, \ldots, n_k$ están ordenados del menor al mayor y $k \geq 2$. Entonces:

Si $n \notin \{n_1, \ldots, n_k\}$, como los eventos de la familia $\{A_1, , \ldots, A_n\}$ son independientes, se tiene:

$P\left(A_{n_1} \cap \cdots \cap A_{n_j}\right) = P\left(A_{n_1}\right) \cdots P\left(A_{n_j}\right)$.

Si $n \in \{n_1, \ldots, n_k\}$, entonces $n_k = n$ y se tiene:

$$P\left(A_{n_1} \cap \cdots \cap A_{n_{k-1}} \cap A_n^c\right)$$

$$= P\left(A_{n_1} \cap \cdots \cap A_{n_{k-1}} - A_{n_1} \cap \cdots \cap A_{n_{k-1}} \cap A_n\right)$$

$$= P\left(A_{n_1} \cap \cdots \cap A_{n_{k-1}}\right) - P\left(A_{n_1} \cap \cdots \cap A_{n_{k-1}} \cap A_n\right)$$

$$= P\left(A_{n_1}\right) \cdots P\left(A_{n_{k-1}}\right) - P\left(A_{n_1}\right) \cdots P\left(A_{n_k-1}\right) P\left(A_n\right)$$

$$= P\left(A_{n_1}\right) \cdots P\left(A_{n_{k-1}}\right)\left[1 - P\left(A_n\right)\right] = P\left(A_{n_1}\right) \cdots P\left(A_{n_{k-1}}\right) P\left(A_n^c\right).$$

Así que los eventos de la familia $\{A_1, , \ldots, A_{n-1}, A_n^c\}$ son independientes.

■

COROLARIO 13.1. *Sea $\{A_1, , \ldots, A_n\}$ una familia de eventos independientes y U un subconjunto del conjunto $\{1, \ldots, n\}$ Para cada $j \in \{1, \ldots, n\}$, definamos:*

$$B_j = \begin{cases} A_j & si\ j \in U \\ A_j^c & si\ j \notin U \end{cases}$$

Entonces, los eventos de la familia $\{B_1, \ldots, B_n\}$ son independientes.

COROLARIO 13.2. *Sean $\{A_\gamma\}_{\gamma\in\Gamma}$ una familia de eventos independientes y T un subconjunto de Γ. Para cada $\gamma \in \Gamma$, definamos:*

$$B_\gamma = \begin{cases} A_\gamma & si\ \gamma \in T \\ A_\gamma^c & si\ \gamma \notin T \end{cases}$$

Entonces, los eventos de la familia $\{B_\gamma\}_{\gamma\in\Gamma}$ son independientes.

TEOREMA 13.1. *Dado un evento A de probabilidad positiva, la función que asigna a cada evento B el número real $P(B \mid A)$, de acuerdo con la definición, es una medida de probabilidad.*

Demostración

La función así definida es claramente no negativa, además:

$P(\Omega \mid A) = \frac{P(A\cap\Omega)}{P(A)} = \frac{P(A)}{P(A)} = 1.$

Finalmente, si $\{B_n\}_{n\in\mathbb{N}}$ es una familia de eventos eventos mutuamente excluyentes, entonces:

$P(\cup_{n=1}^{\infty} B_n \mid A) = \frac{P\left[A\cap(\cup_{n=1}^{\infty} B_n)\right]}{P(A)} = \frac{P\left[\cup_{n=1}^{\infty}(A\cap B_n)\right]}{P(A)} = \frac{\sum_{n=1}^{\infty} P(A\cap B_n)}{P(A)}$

$= \sum_{n=1}^{\infty} \frac{P(A\cap B_n)}{P(A)} = \sum_{n=1}^{\infty} P(B_n \mid A).$

■

La definición de probabilidad condicional implica inmediatamente el siguiente resultado:

PROPOSICIÓN 13.4 (**Regla del producto**). *Sean $A_1, \ldots, A_n$ n eventos tales que $P(A_1 \cap \ldots \cap A_{n-1}) > 0$, entonces:*

$P(\cap_{k=1}^{n} A_k) = P(A_n|A_1 \cap \ldots \cap A_{n-1}) \cdots P(A_2|A_1)P(A_1).$

Combinando la σ-aditividad y la regla del producto se obtiene el siguiente resultado:

PROPOSICIÓN 13.5 (**Regla de la probabilidad total**). *Sean B un evento cualquiera y $A_1, A_2 \ldots$ una colección finita o infinita numerable de eventos de probabilidad positiva, mutuamente excluyentes y tales que $P\left(\bigcup_n A_n\right) = 1$, entonces:*

$P(B) = \sum_n P(B \mid A_n)P(A_n).$

El lema de Borel-Cantelli es un resultado básico en la Teoría de la Probabilidad. Lo vamos a utilizar frecuentemente.

LEMA 13.1. *Sea $\{p_n\}$ una sucesión de números reales en el intervalo $[0,1]$ tal que la serie $\sum_{n=1}^{\infty} p_n$ es divergente, entonces $\lim_{n\to\infty} \prod_{j=1}^{n}(1-p_j) = 0$.*

Por el teorema del valor medio, se tiene, para cada $x \in [0,1)$:

$\ln(1-x) = -\frac{x}{1-\theta x}$, donde $\theta \in (0,1)$.

De manera que, para cualquier $x \in [0,1)$, $\ln(1-x) \leq -x$, es decir, $1-x \leq e^{-x}$, resultado que también es válido para $x = 1$.

En particular, $\prod_{j=1}^{n}(1-p_j) < e^{-\sum_{j=1}^{n} p_n}$, así que $\lim_{n\infty} \prod_{j=1}^{n}(1-p_j) = 0$.

■

TEOREMA 13.2 (**Lema de Borel-Cantelli-1a. parte**). *Sea $A_1, A_2, \ldots$ una sucesión de eventos tales que $\sum_{n=1}^{\infty} P\left(A_n\right) < \infty$ y sea:*

$A = \{\omega \in \Omega : \omega \in A_n$ *para una infinidad de valores de* $n\}$.

Entonces $P(A) = 0$.

Demostración

Para cada $m \in \mathbb{N}$, sea $B_m = \bigcup_{n=m}^{\infty} A_n$. Entonces la sucesión de eventos B_m es decreciente y $A = \bigcap_{m=1}^{\infty} B_m$, así que:

$P(A) = P\left[\bigcap_{m=1}^{\infty} B_m\right] = \text{lím}_{m\rightsquigarrow\infty} P\left[B_m\right]$.

Pero, $P(B_m) = P(\bigcup_{n=m}^{\infty} A_n) \leq \sum_{n=m}^{\infty} P(A_n)$.

Por lo tanto, $P(A) \leq \text{lím}_{m\rightsquigarrow\infty} \sum_{n=m}^{\infty} P(A_n) = 0$.

■

TEOREMA 13.3 (**Lema de Borel-Cantelli-2a. parte**). *Sea* $A_1, A_2, \ldots$ *una sucesión de eventos independientes tales que* $\sum_{n=1}^{\infty} P(A_n) = \infty$ *y sea:*

$A = \{\omega \in \Omega : \omega \in A_n$ *para una infinidad de valores de* $n\}$.

Entonces $P(A) = 1$.

Demostración

Para cada $m \in \mathbb{N}$, sea $B_m = \bigcap_{n=m}^{\infty} A_n^c$. Entonces la sucesión de eventos B_m es creciente y $A^c = \bigcup_{m=1}^{\infty} B_m$, así que:

$P(A^c) = P\left[\bigcup_{m=1}^{\infty} B_m\right] = \text{lím}_{m\to\infty} P\left[B_m\right]$.

Pero, $B_m \subset \bigcap_{n=m}^{m+k} A_n^c$ para cualquier $k \in \mathbb{N}$, así que, $P(B_m) \leq \prod_{n=m}^{m+k} [1 - P(A_n)]$ para cualquier $k \in \mathbb{N}$. Por lo tanto:

$P(B_m) \leq \text{lím}_{k\rightsquigarrow\infty} \prod_{n=m}^{m+k} [1 - P(A_n)] = 0$.

Se concluye entonces que $P(A) = 1 - P(A^c) = 1$.

■

13.1.1. Algunos ejemplos de espacios de probabilidad. Cualquier espacio de medida tal que la medida del total es 1 es un espacio de probabilidad, pero conviene explicitar algunos que son útiles en el desarrollo de la teoría y que nos servirán en algunos ejemplos.

EJEMPLO 13.1. *Si* $A \subset \mathbb{R}$ *es un conjunto Lebesgue medible y su medida es igual a* 1*, entonces, definiendo* $\Omega = A$, $\Im = \{B \cap A : B \subset \mathbb{R}$ *es un conjunto Lebesgue medible*$\}$ *y* P *la medida de Lebesgue restringida a* $\Im$*, entonces* $(\Omega, \Im, P)$ *es un espacio de probabilidad.*

EJEMPLO 13.2. *Si* $A \subset \mathbb{R}^n$ *es un conjunto Lebesgue medible en* $\mathbb{R}^n$ *y su medida es igual a* 1*, entonces, definiendo* $\Omega = A$, $\Im = \{B \cap A : B \subset \mathbb{R}$ *es un conjunto Lebesgue medible en* $\mathbb{R}^n\}$

y P la medida de Lebesgue en $\mathbb{R}^n$ restringida a $\Im$, entonces $(\Omega, \Im, P)$ es un espacio de probabilidad.

EJEMPLO 13.3. *De acuerdo con los resultados del capítulo 5, si $F : \mathbb{R} \mapsto \mathbb{R}$ es una función no decreciente y continua por la derecha, existe una única medida μ_F definida sobre $\mathfrak{B}(\mathbb{R})$ tal que $\mu_F((a, b]) = F(b) - F(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$. Por construcción, la medida μ_F está definida sobre una σ-álgebra más grande que $\mathfrak{B}(\mathbb{R})$, a saber, la σ-álgebra generada por $\mathfrak{B}(\mathbb{R})$ y los subconjuntos de $\mathbb{R}$ contenidos en algún conjunto boreliano de medida μ_F cero. Denotaremos a esta σ-álgebra por $\mathfrak{B}_{\mu_F}(\mathbb{R})$.*

Supongamos ahora que:

$\lim_{x \rightsquigarrow \infty} F(x) - \lim_{x \to -\infty} F(x) = 1.$

Entonces, definiendo $\Omega = \mathbb{R}$, $\Im = \mathfrak{B}_{\mu_F}(\mathbb{R})$ y $P = \mu_F$, $(\Omega, \Im, P)$ es un espacio de probabilidad.

EJEMPLO 13.4. *De la misma manera, si $F : \mathbb{R} \mapsto \mathbb{R}$ es una función no decreciente y continua por la izquierda, existe una única medida μ_F definida sobre $\mathfrak{B}(\mathbb{R})$ tal que $\mu_F([a, b)) = F(b) - F(a)$ para cualquier pareja de números reales, a y b, tales que $a < b$. Nuevamente, denotaremos por $\mathfrak{B}_{\mu_F}$ a la σ-álgebra generada por $\mathfrak{B}(\mathbb{R})$ y los subconjuntos de $\mathbb{R}$ contenidos en algún conjunto boreliano de medida μ_F cero.*

Así que, si se cumple que:

$\lim_{x \rightsquigarrow \infty} F(x) - \lim_{x \to -\infty} F(x) = 1.$

Entonces, definiendo $\Omega = \mathbb{R}$, $\Im = \mathfrak{B}_{\mu_F}(\mathbb{R})$ y $P = \mu_F$, $(\Omega, \Im, P)$ es un espacio de probabilidad.

EJEMPLO 13.5. *De la proposición 3.2 se sigue que si $\Omega = \{\omega_1, \omega_2, \ldots\}$ es un conjunto infinito numerable y $(p_n)_{n \in \mathbb{N}}$ una sucesión de números reales no negativos tales que $\sum_{n=1}^{\infty} p_n = 1$, y para cada $n \in \mathbb{N}$, definimos $p(\omega_n) = p_n$ y, para cualquier subconjunto A de Ω, definimos:*

$P(A) = \sum_{\{\omega \in A\}} p(\omega).$

Entonces P es una medida de probabilidad definida sobre el conjunto potencia de Ω.

EJEMPLO 13.6. *Un ejemplo muy simple es el siguiente:*

Sea $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ un conjunto finito y, $p_1, p_2, \ldots, p_N$ números reales no negativos tales que $\sum_{n=1}^{N} p_n = 1$. Para cada $n \in \{1, 2, \ldots, N\}$, definamos $p(\omega_n) = p_n$ y, para cualquier subconjunto A de Ω, definamos:

$P(A) = \sum_{\{\omega \in A\}} p(\omega).$

Entonces P es una medida de probabilidad definida sobre el conjunto potencia de Ω.

EJEMPLO 13.7. *Un caso particular del ejemplo anterior es el siguiente:*

Sea $\mathbb{F} = \{x_1, x_2, \ldots, x_N\}$ un conjunto finito,

$\Omega_n = \{(y_1, y_2, \ldots, y_n) : y_j \in \{x_1, x_2, \ldots, x_N\}$ *para cualquier* $j \in \{1, 2, \ldots, n\}\}$, *donde* $n \in \mathbb{N}$*, y, para cada* $j \in \{1, 2, \ldots, n\}$*,* $p_1^{(j)}, p_2^{(j)}, \ldots, p_N^{(j)}$ *números reales no negativos tales que* $\sum_{k=1}^{N} p_k^{(j)} = 1$*. Para cada pareja* $j \in \{1, 2, \ldots, n\}$ *y* $k \in \{1, 2, \ldots, N\}$*, definamos* $q_j(x_k) = p_k^{(j)}$*. Definamos también* $p_n : \Omega_n \to \mathbb{R}$ *de la siguiente manera:*

$p_n(y_1, y_2, \ldots, y_n) = \prod_{j=1}^{n} q_j(y_j)$.

Finalmente, para cualquier subconjunto A *de* Ω_n*, definamos:*

$P(A) = \sum_{\{\omega \in A\}} p_n(\omega)$.

Entonces P *es una medida de probabilidad definida sobre el conjunto potencia de* Ω_n*.*

La condición $\sum_{\{\omega \in \Omega_n\}} p_n(\omega) = 1$ *que se requiere, se demuestra a continuación.*

LEMA 13.2. *Sea* $\mathbb{F} = \{x_1, x_2, \ldots, x_N\}$ *un conjunto finito,*

$\Omega_n = \{(y_1, y_2, \ldots, y_n) : y_j \in \{x_1, x_2, \ldots, x_N\}$ *para cualquier* $j \in \{1, 2, \ldots, n\}\}$,

donde $n \in \mathbb{N}$*, y, para cada* $j \in \{1, 2, \ldots, n\}$*,* $p_1^{(j)}, p_2^{(j)}, \ldots, p_N^{(j)}$ *números reales no negativos tales que* $\sum_{k=1}^{N} p_k^{(j)} = 1$*.*

Para cada pareja $j \in \{1, 2, \ldots, n\}$ *y* $k \in \{1, 2, \ldots, N\}$*, definamos:*

$q_j(x_k) = p_k^{(j)}$.

Finalmente, definamos $p_n : \Omega_n \to \mathbb{R}$ *de la siguiente manera:*

$p_n(y_1, y_2, \ldots, y_n) = \prod_{j=1}^{n} q_j(y_j)$.

Entonces:

$\sum_{\{\omega \in \Omega_n\}} p_n(\omega) = 1$.

Demostración

Para cada $n \in \mathbb{N}$, definamos $S(n) = \sum_{\{\omega \in \Omega_n\}} p_n(\omega)$. Entonces, $S(1) = \sum_{k=1}^{N} q_1(x_k) = \sum_{k=1}^{N} p_k^{(1)} = 1$ y $S(n+1) = \sum_{k=1}^{N} q_{n+1}(x_k)\, S(n) = S(n) \sum_{k=1}^{N} p_k^{(n+1)} = S(n)$ para toda $n \in \mathbb{N}$. Así que, por el principio de inducción matemática, $S(n) = 1$ para cualquier $n \in \mathbb{N}$.

■

COROLARIO 13.3. *Sea* $\Omega_n = \{(s_1, s_2, \ldots, s_n) : s_j \in \{0, 1\}$ *para cualquier* $j \in \{1, 2, \ldots, n\}\}$*, donde* $n \in \mathbb{N}$*. Sea* $p \in [0, 1]$ *y, para cada* $\omega = (s_1, \ldots, s_n) \in \Omega_n$*, definamos:*

$p_n(\omega) = \prod_{j=1}^{n} [p s_j + (1 - p)(1 - s_j)]$.

Entonces, $\sum_{\{\omega \in \Omega_n\}} p_n(\omega) = 1$*.*

13.2. Variables Aleatorias

Desde el inicio del Cálculo de Probabilidades se trataba ya con cantidades que podían tomar diferentes valores, cada uno con una determinada probabilidad. En un principio se referían a la ganancia que un jugador podía obtener en un juego de azar. Más tarde se trataron otro tipo de problemas donde también intervenían cantidades que podían tomar distintos valores. A principios del siglo XX se hablaba simplemente de cantidades o variables. En su libro de 1925, Paul Lévy se refería a esas cantidades como variables eventuales. Markov se refería a ellas como variables aleatorias. Una vez que se formula axiomaticamente la Teoría de la Probabilidad, las variables aleatorias quedan identificadas con las funciones medibles.

Definición 13.5. *Llamaremos variable aleatoria real a cualquier función medible de* $(\Omega, \Im)$ *en* $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. *Una variable aleatoria con valores en el conjunto de números reales extendido será una función medible de* $(\Omega, \Im)$ *en* $\left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$. *Un vector aleatorio real será cualquier función medible de* $(\Omega, \Im)$ *en* $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$.

Obviamente, una variable aleatoria real puede considerarse también como una función de $(\Omega, \Im)$ en $\left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$, y esta función es medible.

Dado un conjunto finito de variables aleatorias, $X_1, X_2, \ldots, X_n$, con valores en $\overline{\mathbb{R}}^n$, por la proposición 7.7, la función $(X_1, X_2, \ldots, X_n) : \Omega \to \overline{\mathbb{R}}^n$ definida por:

$$(X_1, X_2, \ldots, X_n)(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$$

es medible. A una función así definida la llamaremos vector aleatorio con valores en $\overline{\mathbb{R}}^n$. También usaremos la notación $\overline{X}$ para un vector aleatorio $(X_1, X_2, \ldots, X_n)$.

A menos que se indique otra cosa, una variable aleatoria (resp. vector aleatorio con n componentes) será considerada con valores en $\overline{\mathbb{R}}$ (resp. con valores en $\overline{\mathbb{R}}^n$).

En lo que se refiere a la convergencia de una sucesión de variables aleatorias, hay algunos cambios en la terminología. En el contexto de la Teoría de la Probabilidad, la convergencia casi en todas partes será denominada **convergencia casi segura** y a la convergencia en medida la llamaremos **convergencia en probabilidad**.

Dada una variable aleatoria $X : \Omega \to \overline{\mathbb{R}}$ y $B \in \mathfrak{B}\left(\overline{\mathbb{R}}\right)$, la notación $[X \in B]$ será una manera abreviada de representar al conjunto $\{\omega \in \Omega : X(\omega) \in B\}$. Si $a, b \in \overline{\mathbb{R}}$, las notaciones $[a < X < b]$, $[a \leq X \leq b]$, $[a \leq X < b]$, $[a < X \leq b]$, $[X < b]$, $[X \leq b]$, $[X > a]$ y $[X \geq a]$ se entenderán en un sentido similar. Para el caso de un vector aleatorio utilizaremos una notación análoga.

Si $(X_1, X_2, \ldots, X_n)$ es un vector aleatorio con valores en $\overline{\mathbb{R}}^n$ y $B_1, B_2, \ldots, B_n$ son conjuntos borelianos en $\overline{\mathbb{R}}$, denotaremos por $[X_1 \in B, X_2 \in B_2, \ldots, X_n \in B_n]$ al conjunto $\bigcap_{k=1}^n [X_k \in B_n]$.

Definición 13.6. *Sea* X *una variable aleatoria con valores en* $\overline{\mathbb{R}}$. *La proyección de* P *bajo* X *será denotada por* μ_X *y la llamaremos la distribución de la variable aleatoria* X. *Si*

$(X_1, X_2, \ldots, X_n)$ es un vector aleatorio, la proyección de P bajo $(X_1, X_2, \ldots, X_n)$ será denotada por $\mu_{X_1,X_2,\ldots,X_n}$ y la llamaremos la distribución del vector aleatorio $(X_1, X_2, \ldots, X_n)$.

De acuerdo con la proposición 8.15 y el corolario 8.8, si X es una variable aleatoria con valores en $\overline{\mathbb{R}}$, μ_X su distribución y $f : \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right), \mu_X\right) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ una función medible, no negativa o integrable, se tiene:

$$\int_{\nleqq} f(X)\, dP = \int_{\overline{\overline{\mathbb{R}}}} f d\mu_X.$$

Si X es una variable aleatoria real y la consideramos como una función de $(\Omega, \Im)$ en $\left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$, entonces X es una variable aleatoria con valores en el conjunto de números reales extendido y su distribución μ_X, aunque está definida sobre $\mathfrak{B}\left(\overline{\mathbb{R}}\right)$, está concentrada en $\mathbb{R}$ ya que $\mu_X(\{-\infty, \infty\}) = 0$.

De la misma manera, si $(X_1, X_2, \ldots, X_n)$ es una vector aleatorio con valores en $\overline{\mathbb{R}}^n$, $\mu_{X_1,X_2,\ldots,X_n}$ su distribución y $f : \left(\overline{\mathbb{R}}^n, \mathfrak{B}\left(\overline{\mathbb{R}}^n\right), \mu_{X_1,X_2,\ldots,X_n}\right) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ una función medible, no negativa o integrable, se tiene:

$$\int_{\nleqq} f(X_1, X_2, \ldots, X_n)\, dP = \int_{\overline{\overline{\mathbb{R}}}} f d\mu_{X_1,X_2,\ldots,X_n}.$$

De manera general, lo que nos interesa de una variable aleatoria, o de un vector aleatorio, es su distribución ya que ésta nos da todas las probabilidades de la forma $P[X \in B]$ (resp. $P[(X_1, X_2, \ldots, X_n) \in B]$, donde B es cualquier conjunto boreliano de $\overline{\mathbb{R}}$ (resp. $\overline{\mathbb{R}}^n$).

Definición 13.7 (**$\boldsymbol{\sigma}$ álgebra generada por una familia de funciones**). *Sea $(\mathbb{E}, \mathfrak{E})$ un espacio medible y $\mathbb{F}$ un conjunto cualquiera. Dada una colección de funciones:*

$$\mathcal{H} = \{f_\gamma : \mathbb{F} \to (\mathbb{E}, \mathfrak{E}) : \gamma \in \Gamma\},$$

donde Γ es un conjunto de índices cualquiera, se define la σ-álgebra generada por $\mathcal{H}$ como la más pequeña σ-álgebra de subconjuntos de $\mathbb{F}$ tal que toda función $f \in \mathcal{H}$ es medible. Denotaremos a esta σ-álgebra por por $\sigma(\mathcal{H})$ o por $\sigma(\{f_\gamma : \gamma \in \Gamma\})$.

Obsérvese que si $f : \mathbb{F} \to (\mathbb{E}, \mathfrak{E})$ es cualquier función, la familia de conjuntos $\{f^{-1}(B) : B \in \mathfrak{E}\}$ es una σ-álgebra de subconjuntos de $\mathbb{F}$. Sin embargo, si $\{f_\gamma : \mathbb{F} \to (\mathbb{E}, \mathfrak{E}) : \gamma \in \Gamma\}$ es una colección de funciones, la familia de conjuntos $\left\{f_\gamma^{-1}(B) : \gamma \in \Gamma \text{ y } B \in \mathfrak{E}\right\}$ no es, en general, una σ-álgebra, pero la σ-álgebra generada por esa familia de conjuntos es la σ-álgebra generada por la familia de funciones $\{f_\gamma : \mathbb{F} \to (\mathbb{E}, \mathfrak{E}) : \gamma \in \Gamma\}$.

Por otra parte, si $f_1, f_2, \ldots, f_n$ son funciones de $\mathbb{F}$ en $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ y definimos $f : \mathbb{F} \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}^n\right)\right)$ mediante la relación:

$$f(x) = (f_1(x), f_2(x), \ldots, f_n(x)),$$

sabemos que si $\Im$ es una σ-álgebra de subconjuntos de $\mathbb{F}$, entonces f es $\Im$-medible si y sólo si f_k es $\Im$-medible para cualquier $k \in \{1, 2, \ldots, n\}$. Por lo tanto:

$\left\{f^{-1}(B) : B \in \mathfrak{B}\left(\overline{\mathbb{R}}^n\right)\right\} = \sigma(f) = \sigma(\{f_1, f_2, \dots, f_n\})$.

PROPOSICIÓN 13.6. *Sean $Y_1, \dots, Y_n$ n variables aleatorias con valores en $\overline{\mathbb{R}}$ y $Z : \Omega \mapsto \overline{\mathbb{R}}$ una función $\sigma(Y_1, \dots, Y_n)$-medible. Entonces, existe una función boreliana $h : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ tal que $Z = h(Y_1, \dots, Y_n)$.*

Demostración

Si $Z = I_E$, donde $E \in \sigma(Y_1, \dots, Y_n)$, entonces existe un boreliano $B \subset \overline{\mathbb{R}}^n$ tal que $Z = I_B(Y_1, \dots, Y_n)$. Por lo tanto, se tiene el resultado para el caso de una función simple Z $\sigma(Y_1, \dots, Y_n)$-medible.

Si Z es una variable aleatoria no negativa, sea $(Z_n)_{n\in\mathbb{N}}$ una sucesión no decreciente de funciones simples no negativas tales que $Z = \lim_{n\rightsquigarrow\infty} Z_n$ y, para cada $n \in \mathbb{N}$, sea $h_n : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ una función boreliana no negativa tal que $Z_n = h_n(Y_1, \dots, Y_n)$.

Sea $D = \left\{x \in \overline{\mathbb{R}}^n : \lim_{n\rightsquigarrow\infty} h_n(x) \text{ existe}\right\}$. Entonces D es un conjunto boreliano de $\overline{\mathbb{R}}^n$ y contiene a la imagen de Ω bajo la función $(Y_1, \dots, Y_n)$. Definamos la función $h : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ de la siguiente manera:

$$h(x) = \begin{cases} \lim_{n\rightsquigarrow\infty} h_n(x) & \text{si } x \in D \\ 0 & \text{en otro caso} \end{cases}$$

h es entonces una función boreliana y $Z = h(Y_1, \dots, Y_n)$.

■

COROLARIO 13.4. *Sean $Y_1, \dots, Y_n$ n variables aleatorias con valores en $\mathbb{R}$ y $Z : \Omega \mapsto \mathbb{R}$ una función $\sigma(Y_1, \dots, Y_n)$-medible. Entonces, existe una función boreliana $h : \mathbb{R}^n \mapsto \mathbb{R}$ tal que $Z = h(Y_1, \dots, Y_n)$.*

13.3. Independencia de variables aleatorias

DEFINICIÓN 13.8 (**Independencia de variables aleatorias**). *Diremos que las variables aleatorias de una familia no vacía cualquiera $\{X_\gamma\}$, finita o infinita, son independientes, si dada cualquier subcolección finita de ellas, $X_{\gamma_1}, \dots, X_{\gamma_n}$ y cualquier colección de subconjuntos borelianos de $\overline{\mathbb{R}}$, $A_1, \dots, A_n$, donde $n \in \mathbb{N}$, se tiene:*

$$P\left[X_{\gamma_1} \in A_1, \dots, X_n \in X_{\gamma_n}\right] = P\left[X_{\gamma_1} \in A_1\right] \cdots P\left[X_{\gamma_n} \in A_n\right].$$

PROPOSICIÓN 13.7. *n variables aleatorias $X_1, \dots, X_n$, son independientes si y sólo si para cualquier colección de subconjuntos borelianos de $\overline{\mathbb{R}}$, $A_1, \dots, A_n$, se tiene:*

$$P[X_1 \in A_1, \dots, X_n \in A_n] = P[X_1 \in A_1] \cdots P[X_n \in A_n].$$

Demostración

Si las n variables aleatorias son independientes, la relación se sigue de la definición.

Supongamos ahora que para cualquier colección de subconjuntos borelianos de $\overline{\mathbb{R}}$, $A_1, \ldots, A_n$, se tiene:

$$P[X_1 \in A_1, \ldots, X_n \in A_n] = P[X_1 \in A_1] \cdots P[X_n \in A_n].$$

Sean $\{X_{i_1}, \ldots, X_{i_k}\}$ un subconjunto de la familia $\{X_1, \ldots, X_n\}$ y $A_{i_1}, \ldots, A_{i_k}$ subconjuntos borelianos de $\overline{\mathbb{R}}$. Para $j \in \{1, \ldots, n\} - \{i_1, \ldots, i_k\}$, definamos $A_j = \overline{\mathbb{R}}$, entonces:

$$P[X_{i_1} \in A_{i_1}, \ldots, X_{i_k} \in A_{i_k}] = P[X_1 \in A_1, \ldots, X_n \in A_n]$$

$$= P[X_1 \in A_1] \cdots P[X_n \in A_n] = P[X_{i_1} \in A_{i_1}] \cdots P[X_{i_k} \in A_{i_k}].$$

Así que $X_1, \ldots, X_n$ son independientes.

■

Corolario 13.5. *Las variables aleatorias de una familia infinita numerable, $\{X_1, X_2, \ldots\}$, son independientes si y sólo si para cualquier $n \in \mathbb{N}$ y cualquier colección de subconjuntos borelianos en $\overline{\mathbb{R}}$, $A_1, \ldots, A_n$, se tiene:*

$$P[X_1 \in A_1, \ldots, X_n \in A_n] = P[X_1 \in A_1] \cdots P[X_n \in A_n].$$

Teorema 13.4. *Sean $X_1, \ldots, X_n$ n variables aleatorias independientes y $f_1, \ldots, f_n$ n funciones borelianas de $\overline{\mathbb{R}}$ en $\overline{\mathbb{R}}$. Entonces las variables aleatorias $f_1(X_1), \ldots, f_n(X_n)$ son independientes.*

Demostración

Sean $A_1, \ldots, A_n$ subconjuntos borelianos de $\overline{\mathbb{R}}$, entonces $f_1^{-1}(A_1), \ldots, f_n^{-1}(A_n)$ son también subconjuntos borelianos de $\overline{\mathbb{R}}$, así que:

$$P[f_1(X_1) \in A_1, \ldots, f_n(X_n) \in A_n] = P\left[X_1 \in f_1^{-1}(A_1), \ldots, X_n \in f_n^{-1}(A_n)\right]$$

$$= P\left[X_1 \in f_1^{-1}(A_1)\right] \cdots P[X_n \in f_n^{-1}(A_n)] = P[f_1(X_1) \in A_1] \cdots P[f_n(X_n) \in A_n].$$

■

Teorema 13.5. *Sean $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m}$ $n+m$ variables aleatorias independientes y $f : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ y $g : \overline{\mathbb{R}}^m \mapsto \overline{\mathbb{R}}$ dos funciones borelianas. Entonces, las variables aleatorias $f(X_1, \ldots, X_n)$ y $g(X_{n+1}, \ldots, X_{n+m})$ son independientes.*

Demostración

Sea $\mathcal{G}$ la familia de subconjuntos $B \subset \overline{\mathbb{R}}^m$ tales que:

$$P[(X_1, \ldots, X_n) \in A_1 \times A_2 \times \ldots \times A_n, (X_{n+1}, \ldots, X_{n+m}) \in B]$$

$$= P[(X_1, \ldots, X_n) \in A_1 \times A_2 \times \ldots \times A_n] \, P[(X_{n+1}, \ldots, X_{n+m}) \in B]$$

para cualquier colección $A_1, \ldots, A_n$ de boreliano de $\overline{\mathbb{R}}$.

$\mathcal{G}$ es entonces un d-sistema que contiene a los borelianos de $\overline{\mathbb{R}}^m$ de la forma $B_1 \times B_2 \times \ldots \times B_n$, donde $B_1, \ldots, B_n$ son borelianos de $\overline{\mathbb{R}}$, así que, por el teorema de clases monótonas para pi-sistemas, $\mathcal{G}$ contiene a todos los borelianos de $\overline{\mathbb{R}}^m$.

Sea $\mathcal{H}$ la familia de subconjuntos $A \subset \overline{\mathbb{R}}^n$ tales que:

$P\left[(X_1, \ldots, X_n) \in A, (X_{n+1}, \ldots, X_{n+m}) \in B\right]$

$= P\left[(X_1, \ldots, X_n) \in A\right] P\left[(X_{n+1}, \ldots, X_{n+m}) \in B\right]$

para cualquier boreliano B de $\overline{\mathbb{R}}^m$.

$\mathcal{H}$ es entonces un d-sistema que contiene a los borelianos de $\overline{\mathbb{R}}^m$ de la forma $A_1 \times A_2 \times \ldots \times A_n$, donde $A_1, \ldots, A_n$ son borelianos de $\overline{\mathbb{R}}$, así que, por el teorema de clases monótonas para pi-sistemas, $\mathcal{H}$ contiene a todos los borelianos de $\overline{\mathbb{R}}^n$.

Por lo tanto:

$P\left[(X_1, \ldots, X_n) \in A, (X_{n+1}, \ldots, X_{n+m}) \in B\right]$

$= P\left[(X_1, \ldots, X_n) \in A\right] P\left[(X_{n+1}, \ldots, X_{n+m}) \in B\right]$

para cualquier pareja de borelianos A y B de $\overline{\mathbb{R}}^n$ y $\overline{\mathbb{R}}^m$, respectivamente.

Sean C y D subconjuntos borelianos de $\overline{\mathbb{R}}$, entonces $f^{-1}(C)$ y $g^{-1}(D)$ son subconjuntos borelianos de $\overline{\mathbb{R}}^n$ y $\overline{\mathbb{R}}^m$, respectivamente. Por lo tanto:

$P\left[f(X_1, \ldots, X_n) \in C, g(X_{n+1}, \ldots, X_{n+m}) \in D\right]$

$= P\left[(X_1, \ldots, X_n) \in f^{-1}(C), (X_{n+1}, \ldots, X_{n+m}) \in g^{-1}(D)\right]$

$= P\left[(X_1, \ldots, X_n) \in f^{-1}(C)\right] P\left[(X_{n+1}, \ldots, X_{n+m}) \in g^{-1}(D)\right]$

$= P\left[f(X_1, \ldots, X_n) \in C\right] P\left[g(X_{n+1}, \ldots, X_{n+m}) \in D\right].$

■

13.4. Funciones de distribución

Definición 13.9 (**Función de distribución**). *Si X es una variable aleatoria real, la función $F_X : \mathbb{R} \mapsto \mathbb{R}$, definida por $F_X(x) = P[X \leq x]$, es llamada la función de distribución de X.*

Proposición 13.8. *Sea X una variable aleatoria real y F_X su función de distribución, entonces:*

(i) *F_X es una función no decreciente y continua por la derecha.*
(ii) $\lim_{x \rightsquigarrow \infty} F_X(x) = 1$

(iii) $\lim_{x \to -\infty} F_X(x) = 0$
(iv) $F_X(x-) = P[X < x]$ *para cualquier* $x \in \mathbb{R}$.

Demostración

i. Sea $(x_n)_{n\in\mathbb{N}}$ una sucesión decreciente tal que $\lim_{n\rightsquigarrow\infty} x_n = x$, entonces:

$F_X(x+) = \lim_{n\rightsquigarrow\infty} F_X(x_n) = \lim_{n\to\infty} P[X \leq x_n] = P\left(\bigcap_{n=1}^{\infty} [X \leq x_n]\right)$

$= P[X \leq x] = F_X(x)$.

ii. Sea (x_n) una sucesión creciente tal que $\lim_{n\rightsquigarrow\infty} x_n = \infty$, entonces:

$\lim_{n\rightsquigarrow\infty} F_X(x_n) = \lim_{n\rightsquigarrow\infty} P[X \leq x_n] = P\left(\bigcup_{n=1}^{\infty} [X \leq x_n]\right)$

$= P[\Omega] = 1$.

iii. Sea (x_n) una sucesión decreciente tal que $\lim_{n\rightsquigarrow\infty} x_n = -\infty$, entonces:

$\lim_{n\rightsquigarrow\infty} F_X(x_n) = \lim_{n\rightsquigarrow\infty} P[X \leq x_n] = P\left(\bigcap_{n=1}^{\infty} [X \leq x_n]\right) = P[\emptyset] = 0$.

iv. Sea $x \in \mathbb{R}$ y (x_n) una sucesión creciente tal que $\lim_{n\rightsquigarrow\infty} x_n = x$, entonces:

$F_X(x-) = \lim_{n\rightsquigarrow\infty} F_X(x_n) = \lim_{n\to\infty} P[X \leq x_n] = P\left(\bigcup_{n=1}^{\infty} [X \leq x_n]\right) = P[X < x]$.

■

Obsérvese que si X es una variable aleatoria real y F_X su función de distribución, entonces μ_X es la medida generada por F_X.

Sea X una variable aleatoria real y definamos:

$D = \{x \in \mathbb{R} : P[X = x] > 0\}$,

$C = \{x \in \mathbb{R} : P[X = x] = 0\}$,

$p = P[X \in D]$.

Clasificaremos a las variables aleatorias de acuerdo al valor de p. Si $p = 0$, diremos que la variable aleatoria es continua, si $p = 1$ diremos que es discreta. Cuando $0 < p < 1$, se tiene $P[X \in D] > 0$ y $P[X \in C] > 0$, de manera que se puede decir que, en ese caso, la variable aleatoria tiene una parte discreta (la que corresponde al conjunto D) y una parte continua (la que corresponde al conjunto C).

Un subconjunto importante del conjunto de variables aleatorias continuas está formado por las variables aleatorias reales X cuya distribución μ_X es absolutamente continua con respecto a la medida de Lebesgue restringida a $\mathfrak{B}\left(\overline{\mathbb{R}}\right)$. En este caso μ_X se puede extender de manera única a $\mathfrak{L}\left(\overline{\mathbb{R}}\right)$ ya que si $E \in \mathfrak{L}(\mathbb{R})$, consideremos $B \in \mathfrak{B}(\mathbb{R})$ y $C \in \mathfrak{L}(\mathbb{R})$, de medida de Lebesgue cero, tales que $E = B \cup C$; tomemos entonces $F \in \mathfrak{B}(\mathbb{R})$ de medida de Lebesgue cero tal que $C \subset F$. Como μ_X es absolutamente continua con respecto a la medida de Lebesgue restringida a $\mathfrak{B}\left(\overline{\mathbb{R}}\right)$, entonces, $\mu_X(F) = 0$. Así que podemos extender, de manera única, μ_X a $\mathfrak{L}\left(\overline{\mathbb{R}}\right)$ definiendo $\mu_X(G) = 0$ para cualquier conjunto $G \in \mathfrak{L}\left(\overline{\mathbb{R}}\right)$ de medida de Lebesgue cero. Diremos que una variable aleatoria de este tipo es absolutamente continua.

Definición 13.10. *Si X es una variable aleatoria discreta, llamaremos función de densidad de X, y la denotaremos por f_X, a la función $f_X : \mathbb{R} \to [0,1]$ definida mediante la relación:*

$f_X(x) = P[X = x]$.

Definición 13.11. *Si X es una variable aleatoria absolutamente continua, llamaremos función de densidad de X, y la denotaremos por f_X, a cualquier función medible $f : (\mathbb{R}, \mathfrak{L}(\mathbb{R})) \to \mathbb{R}^+$ tal que:*

$\mu_X(B) = \int_B f d\lambda$

para cualquier conjunto $B \in \mathfrak{L}\left(\overline{\mathbb{R}}\right)$.

13.4.1. Algunos ejemplos de distribuciones. Cualquier medida de probabilidad sobre $\left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una distribución de probabilidad, es decir, una distribución de alguna variable aleatoria. La demostración de esta afirmación es inmediata, ya que si μ es una medida de probabilidad sobre $\left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$, entonces podemos tomar como espacio de probabilidad a la terna $\left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right), \mu\right)$. Entonces la variable aleatoria $X : \Omega \to \mathbb{R}$ definida por $X(\omega) = \omega$ tiene como distribución a la medida μ. Sin embargo, es importante explicitar algunas de ellas, ya sea por su interés histórico o por presentarse con frecuencia en diferentes problemas. Así que a continuación haremos un listado de algunas distribuciones discretas y algunas absolutamente continuas, dando el nombre de la distribución y la función de densidad que la determina.

Distribución Bernoulli con parámetro p, donde $p \in [0,1]$.

$$f(x) = \begin{cases} p & \text{si } x = 1 \\ 1-p & \text{si } x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

Distribución binomial con parámetros n y p, donde $n \in \mathbb{N}$ y $p \in [0,1]$.

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x \in \{0, 1, \ldots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

Distribución geométrica con parámetro p, donde $p \in [0,1]$.

$$f(x) = \begin{cases} p(1-p)^x & \text{si } x \in \{0, 1, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

Distribución binomial negativa con parámetros n y p, donde $n \in \mathbb{N}$ y $p \in [0,1]$.

$$f(x) = \begin{cases} \binom{n+x-1}{x} p^n (1-p)^x & \text{si } x \in \{0, 1, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

Distribución Poisson con parámetro λ, donde λ es un número real positivo.

$$f(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{si } x \in \{0, 1, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

Distribución hipergeométrica con parámetros r, s y n, donde $r, s, n \in \mathbb{N}$ y $n \leq r + s$.

$$f(x) = \begin{cases} \frac{\binom{r}{x}\binom{s}{n-x}}{\binom{r+s}{n}} & \text{si } x \in \{0, 1, \ldots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

Distribución uniforme discreta en el conjunto A, donde A es un conjunto finito de números reales.

$$f_X(x) = \begin{cases} \frac{1}{N} & \text{si } x \in A \\ 0 & \text{en otro caso} \end{cases}$$

donde N es el número de elementos de A.

Distribución uniforme en el conjunto A, donde A es un conjunto Lebesgue medible de medida positiva.

$$f_X(x) = \begin{cases} \frac{1}{\lambda(A)} & \text{si } x \in A \\ 0 & \text{en otro caso} \end{cases}$$

Distribución normal con parámetros μ y σ^2, donde $\mu, \sigma \in \mathbb{R}$ y $\sigma > 0$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Distribución normal estándar.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Distribución exponencial con parámetro λ, donde λ es un número real positivo.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

distribución gama con parámetros α y λ, donde α y λ son números reales positivos.

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

donde $\Gamma : (0, \infty) \to \mathbb{R}$ es la función gama, la cual está definida por: $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

Si X es una variable aleatoria real, μ_X su distribución y F_X su función de distribución, entonces, de acuerdo con el teorema 1, si definimos $c_X : (0,1) \mapsto \mathbb{R}$ mediante la relación $c_X(t) = \text{ínf}\{x \in \mathbb{R} : F_X(x) > t\}$, se tiene:

$$\mu_X(B) = \lambda\left(c_X^{-1}(B)\right)$$

para cualquier conjunto $B \in \mathfrak{B}(\mathbb{R})$, donde λ es la medida de Lebesgue en el intervalo $(0,1)$.

Sea $U : (\Omega, \Im) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ una variable aleatoria con distribución uniforme en el intervalo $(0,1)$, entonces μ_U coincide con la medida de Lebesgue en el intervalo $(0,1)$ y $\mu_U\left(\overline{\mathbb{R}} - (0,1)\right) = 0$. Así que, si $f : \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right), \mu_U\right) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una función medible, no negativa o integrable, se tiene:

$$\int_{\nleq} f(U)\, dP = \int_{\overline{\mathbb{R}}} f d\mu_U = \int_0^1 f(t)\, dt.$$

Por lo tanto, si $B \in \mathfrak{B}(\mathbb{R})$, se tiene:

$$\mu_X(B) = \lambda\left(c_X^{-1}(B)\right) = \int_0^1 (I_B \circ c_X)(t)\, dt = \int_{\nleq} (I_B \circ c_X)(U)\, dP$$

$$= \int_{\nleq} I_B(c_X(U))\, dP = P[c_X(U) \in B] = \mu_{c_X(U)}(B).$$

Así que la distribución de $c_X(U)$ coincide con la distribución de X.

Este resultado se puede demostrar directamente sin mucha dificultad ya que P es una medida finita.

Definamos $d_X : (0,1) \mapsto \mathbb{R}$ mediante la relación $d_X(t) = \text{ínf}\{x \in \mathbb{R} : F_X(x) \geq t\}$.

LEMA 13.3. *Sea X una variable aleatoria real. Entonces el conjunto* $\{t \in (0,1) : c_X(t) \neq d_X(t)\}$ *es, a lo más, infinito numerable.*

Demostración

Definamos $\mathcal{C}$ como el conjunto de puntos $t \in (0,1)$ para los cuales existe un intervalo de longitud positiva en el cual F_X es constante e igual a t.

Si $t \in (0,1)$ y $t \in \mathcal{C}$, entonces d_X es discontinua en t, así que $\mathcal{C}$ está contenido en el conjunto de puntos donde d_X es discontinua. Por otra parte, al ser d_X una función no decreciente, el conjunto de puntos en los cuales d_X es discontinua es finito o infinito numerable. Por lo tanto, $\mathcal{C}$ es un conjunto finito o infinito numerable.

Si $t \in (0,1)$ y $t \notin \mathcal{C}$, entonces $c_X(t) = d_X(t)$, así que, si $c_X(t) \neq d_X(t)$, entonces $t \in \mathcal{C}$.

Por lo tanto:

$$\{t \in (0,1) : c_X(t) \neq d_X(t)\} \subset \mathcal{C}.$$

lo cual prueba el resultado.

■

TEOREMA 13.6. *Sean X una variable aleatoria real con función de distribución F_X y U una variable aleatoria con distribución uniforme en el intervalo $(0,1)$. Entonces la función de distribución de la variable aleatoria $d_X(U)$ es F_X.*

Demostración

Como F_X es continua por la derecha, se tiene $F_X\left(d_X(t)\right) \geq t$.

Tomemos $z \in \mathbb{R}$.

Si $d_X(U(\omega)) \leq z$, entonces $U(\omega) \leq F_X\left(d_X(U(\omega))\right) \leq F_X(z)$.

Si $U(\omega) \leq F_X(z)$, entonces $d_X\left(U(\omega)\right) = \inf\left\{s \in \mathbb{R} : F_X(s) \geq U(\omega)\right\} \leq z$

Por lo tanto, se tiene:

$P\left[d_X(U) \leq z\right] = P\left[U \leq F_X(z)\right] = F_X(z)$.

∎

COROLARIO 13.6. *Sea X una variable aleatoria con función de distribución F_X y U una variable aleatoria con distribución uniforme en el intervalo $(0,1)$. Entonces la función de distribución de la variable aleatoria $c_X(U)$ es F_X.*

Demostración

Como el conjunto $B = \{t \in (0,1) : c_X(t) \neq d_X(t)\}$ es, a lo más, infinito numerable, $P\left[U \in B\right] = 0$, así que $P\left[c_X(U) = d_X(U)\right] = 1$. Por lo tanto:

$P\left[c_X(U) \leq z\right] = P\left[d_X(U) \leq z\right] = F_X(z)$

para cualquier $z \in \mathbb{R}$.

∎

13.5. Funciones de distribución conjuntas

DEFINICIÓN 13.12 (**Función de distribución conjunta**). *Sean $X_1, \ldots, X_n$ n variables aleatorias reales. La función $F_{X_1,\ldots,X_n} : \mathbb{R}^n \mapsto \mathbb{R}$, definida por:*

$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P\left[X_1 \leq x_1, \ldots, X_n \leq x_n\right]$

será llamada la función de distribución conjunta de $X_1, \ldots, X_n$.

TEOREMA 13.7. *Sean $X_1, \ldots, X_n$ n variables aleatorias y sea $F_{X_1,\ldots,X_n}$ su función de distribución conjunta, entonces, para cada:*

$(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \in \mathbb{R}^{n-1}$, *se tiene:*

(i) *La función* $x \mapsto F_{X_1,\ldots,X_n}(x_1,\ldots,x_{j-1},x,x_{j+1},\ldots,x_n)$*, definida sobre* $\mathbb{R}$*, es no decreciente y continua por la derecha.*

(ii) $\lim_{x\rightsquigarrow\infty} F_{X_1,\ldots,X_n}(x_1,\ldots,x_{j-1},x,x_{j+1},\ldots,x_n)$
$= F_{X_1,,\ldots,X_{j-1},X_{j+1},\ldots,X_n}(x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n)$.

(iii) $\lim_{x\to-\infty} F_{X_1,\ldots,X_n}(x_1,\ldots,x_{j-1},x,x_{j+1},\ldots,x_n) = 0$.

Demostración

i. Sean $x \in \mathbb{R}$ y $(y_m)_{m\in\mathbb{N}}$ una sucesión monótona decreciente tal que $\lim_{n\rightsquigarrow\infty} y_m = x$, entonces:

$\lim_{m\rightsquigarrow\infty} F_{X_1,\ldots,X_n}(x_1,\ldots,x_{j-1},y_m,x_{j+1},\ldots,x_n)$

$= \lim_{m\to\infty} P\left[X_1 \leq x_1,\ldots,X_{j-1} \leq x_{j-1}, X_j \leq y_m, X_{j+1} \leq x_{j+1},\ldots,X_n \leq x_n\right]$

$= P\left(\bigcap_{m=1}^{\infty}\left[X_1 \leq x_1,\ldots,X_{j-1} \leq x_{j-1}, X_j \leq y_m, X_{j+1} \leq x_{j+1},\ldots,X_n \leq x_n\right]\right)$

$= P\left[X_1 \leq x_1,\ldots,X_{j-1} \leq x_{j-1}, X_j \leq x, X_{j+1} \leq x_{j+1},\ldots,X_n \leq x_n\right]$

$= F_{X_1,\ldots,X_n}(x_1,\ldots,x_{j-1},x,x_{j+1},\ldots,x_n)$.

ii. Sea $(y_m)_{m\in\mathbb{N}}$ una sucesión creciente tal que $\lim_{n\rightsquigarrow\infty} y_m = \infty$, entonces:

$\lim_{m\to\infty} P\left[X_1 \leq x_1,\ldots,X_{j-1} \leq x_{j-1}, X_j \leq y_m, X_{j+1} \leq x_{j+1},\ldots,X_n \leq x_n\right]$

$= P\left(\bigcup_{m=1}^{\infty}\left[X_1 \leq x_1,\ldots,X_{j-1} \leq x_{j-1}, X_j \leq y_m, X_{j+1} \leq x_{j+1},\ldots,X_n \leq x_n\right]\right)$

$= P\left[X_1 \leq x_1,\ldots,X_{j-1} \leq x_{j-1}, X_j \in \mathbb{R}, X_{j+1} \leq x_{j+1},\ldots,X_n \leq x_n\right]$

$= F_{X_1,,\ldots,X_{j-1},X_{j+1},\ldots,X_n}(x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n)$.

iii. Sea $(y_m)_{m\in\mathbb{N}}$ una sucesión decreciente tal que $\lim_{n\rightsquigarrow\infty} y_m = -\infty$, entonces:

$\lim_{m\to\infty} P\left[X_1 \leq x_1,\ldots,X_{j-1} \leq x_{j-1}, X_j \leq y_m, X_{j+1} \leq x_{j+1},\ldots,X_n \leq x_n\right]$

$= P\left(\bigcap_{m=1}^{\infty}\left[X_1 \leq x_1,\ldots,X_{j-1} \leq x_{j-1}, X_j \leq y_m, X_{j+1} \leq x_{j+1},\ldots,X_n \leq x_n\right]\right)$

$= P\left[\emptyset\right] = 0$.

∎

Las condiciones de la proposición anterior no son suficientes para que una función F sea una función de distribución conjunta. En efecto, consideremos, por ejemplo, la siguiente función:

$$F(x,y) = \begin{cases} 0 & \text{si } x < 0 \text{ ó } y < 0 \\ x+y & \text{si } x+y < 1, x \geq 0, y \geq 0 \\ 1 & \text{si } x+y \geq 1, x \geq 0, y \geq 0 \end{cases}$$

Esta función tiene las propiedades siguientes:

(i) Para cada $y \in \mathbb{R}$, la función $x \mapsto F(x,y)$ es no decreciente y continua por la derecha y $\lim_{x \rightsquigarrow -\infty} F(x,y) = 0$.
(ii) Para cada $x \in \mathbb{R}$, la función $y \mapsto F(x,y)$ es no decreciente y continua por la derecha y $\lim_{y \rightsquigarrow -\infty} F(x,y) = 0$.
(iii) Las funciones $G : \mathbb{R} \mapsto [0,1]$ y $H : \mathbb{R} \mapsto [0,1]$, definidas por $G(y) = \lim_{x \rightsquigarrow \infty} F(x,y)$ y $H(x) = \lim_{y \rightsquigarrow \infty} F(x,y)$, respectivamente, son funciones de distribución en una variable.

Sin embargo, F no es una función de distribución conjunta de alguna pareja de variables aleatorias X, Y. En efecto, si lo fuera, se tendría:

$$P[X \leq x] = \lim_{y \rightsquigarrow \infty} F_{X,Y}(x,y) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$$

$$P[Y \leq y] = \lim_{x \rightsquigarrow \infty} F_{X,Y}(x,y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 & \text{si } y \geq 0 \end{cases}$$

Así que, $P[X = 0] = P[Y = 0] = 1$.

Por lo tanto, se tendría $P[X = 0, Y = 0] = 1$.

Pero, $P[X = 0, Y = 0] \leq F(0,0) = 0$, lo cual es una contradicción.

Una familia de variables aleatorias $X_1, \ldots, X_n$ puede verse como la función de Ω en $\mathbb{R}^n$ que asigna a cada $\omega \in \Omega$ el vector $(X_1(\omega), \ldots, X_n(\omega))$; de esta forma, podemos decir que las variables aleatorias forman un **vector aleatorio** $(X_1, \ldots, X_n)$.

Recordemos que un rectángulo en $\mathbb{R}^n$ es un conjunto de la forma $I_1 \times \cdots \times I_n$, en donde $I_1, \cdots, I_n$ son intervalos en $\mathbb{R}$.

Si $R = I_1 \times \cdots \times I_n$ es un rectángulo en $\mathbb{R}^n$ y $a_1, b_1, \ldots, a_n, b_n$ son los extremos de $I_1, \cdots, I_n$, respectivamente, los intervalos I_k serán llamados los lados del rectángulo y los puntos del conjunto $V_{(a_1,b_1,\ldots,a_n,b_n)} = \{(x_1, \ldots, x_n) : x_k \in \{a_k, b_k\} \text{ para toda } k\}$ serán llamados los vértices del rectángulo.

El rectángulo $(a_1, b_1] \times \cdots \times (a_n, b_n]$ será denotado por $R_{(a_1,b_1,\ldots,a_n,b_n)}$ y $S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}$ denotará al conjunto:

$\{(x_1, \cdots, x_n) : x_i = a_i$ para k índices i y $x_i = b_i$ para el resto de índices$\}$.

Obviamente, $V_{(a_1,b_1,\ldots,a_n,b_n)} = \bigcup_{k=0}^{n} S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}$.

Teorema 13.8. *Sean $X_1, \ldots, X_n$ n variables aleatorias y $(a_1, b_1] \times \cdots \times (a_n, b_n]$ un rectángulo de $\mathbb{R}^n$.*

Entonces, para cualquier evento A se tiene:

$P([a_1 < X_1 \leq b_1, \cdots, a_n < X_n \leq b_n] \cap A)$

$= \sum_{k=0}^{n}(-1)^k \sum_{\left\{(x_1,\cdots,x_n)\in S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}\right\}} P\left([X_1 \leq x_1, \cdots, X_n \leq x_n] \cap A\right).$

Demostración

Para $n = 1$ se tiene:

$P\left([a_1 < X_1 \leq b_1] \cap A\right) = P\left([X_1 \leq b_1] \cap A\right) - P\left([X_1 \leq a_1] \cap A\right).$

Así que la relación se cumple en este caso.

Supongamos que la relación se cumple para cualquier rectángulo $(a_1, b_1] \times \cdots \times (a_k, b_k]$ y cualquier evento A. Entonces, dado cualquier rectángulo $(a_1, b_1] \times \cdots \times (a_{k+1}, b_{k+1}]$ y cualquier evento A se tiene:

$P\left([a_1 < X_1 \leq b_1, \cdots, a_{n+1} < X_{n+1} \leq b_{n+1}] \cap A\right)$

$= P\left([a_1 < X_1 \leq b_1, \cdots, a_n < X_n \leq b_n, X_{n+1} \leq b_{n+1}] \cap A\right)$

$-P\left([a_1 < X_1 \leq b_1, \cdots, a_n < X_n \leq b_n, X_{n+1} \leq a_{n+1}] \cap A\right)$

$= \left(P\left[a_1 < X_1 \leq b_1, \cdots, a_n < X_n \leq b_n\right] \cap A \cap [X_{n+1} \leq b_{n+1}]\right)$

$- \left(P\left[a_1 < X_1 \leq b_1, \cdots, a_n < X_n \leq b_n\right] \cap A \cap [X_{n+1} \leq a_{n+1}]\right)$

$= \sum_{k=0}^{n}(-1)^k \sum_{\left\{(x_1,\cdots,x_n)\in S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}\right\}} P\left([X_1 \leq x_1, \cdots, X_n \leq x_n] \cap A \cap [X_{n+1} \leq b_{n+1}]\right)$

$- \sum_{k=0}^{n}(-1)^k \sum_{\left\{(x_1,\cdots,x_n)\in S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}\right\}} P\left([X_1 \leq x_1, \cdots, X_n \leq x_n] \cap A \cap [X_{n+1} \leq a_{n+1}]\right)$

$= \sum_{k=0}^{n+1}(-1)^k \sum_{\left\{(x_1,\cdots,x_{n+1})\in S^{(k)}_{(a_1,b_1,\ldots,a_{n+1},b_{n+1})}\right\}} P\left([X_1 \leq x_1, \cdots, X_{n+1} \leq x_{n+1}] \cap A\right).$

Así que, por el principio de inducción matemática, la relación se cumple para cualquier $n \in \mathbb{N}$, cualquier rectángulo $(a_1, b_1] \times \cdots \times (a_n, b_n]$ y cualquier evento A.

∎

COROLARIO 13.7. *Sean $X_1, \ldots, X_n$ n variables aleatorias y $(a_1, b_1] \times \cdots \times (a_n, b_n]$ un rectángulo de $\mathbb{R}^n$. Entonces:*

$P\left([a_1 < X_1 \leq b_1, \cdots, a_n < X_n \leq b_n]\right)$

$= \sum_{k=0}^{n}(-1)^k \sum_{\left\{(x_1,\cdots,x_n)\in S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}\right\}} F_{X_1,\ldots,X_n}\left(x_1, \cdots, x_n\right).$

COROLARIO 13.8. *Sean $X_1, \ldots, X_n$ n variables aleatorias y $F_{X_1,\ldots,X_n}$ su función de distribución conjunta, entonces:*

$\sum_{k=0}^{n}(-1)^k \sum_{\left\{(x_1,\cdots,x_n)\in S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}\right\}} F_{X_1,\ldots,X_n}\left(x_1, \cdots, x_n\right) \geq 0$

para cualquier rectángulo $(a_1, b_1] \times \cdots \times (a_n, b_n]$.

Para la continuidad por la derecha, se tiene el siguiente resultado, más general que el enunciado en la proposición 13.7:

TEOREMA 13.9. *Sean $X_1, \ldots, X_n$ n variables aleatorias y $F_{X_1,\ldots,X_n}$ su función de distribución conjunta, entonces:*

$$\lim_{m \rightsquigarrow 0} F_{X_1,\ldots,X_n}\left(x_1 + \delta_1^{(m)}, \cdots, x_n + \delta_n^{(m)}\right) = F_{X_1,\ldots,X_n}(x_1, \cdots, x_n)$$

para cualquier vector $(x_1, \cdots, x_n) \in \mathbb{R}^n$ y cualquier sucesión $\left(\left(\delta_1^{(m)}, \cdots, \delta_n^{(m)}\right)\right)_{m\in\mathbb{N}}$ que converja al vector $\overline{0} \in \mathbb{R}^n$ y tal que $\delta_1^{(m)}, \cdots, \delta_n^{(m)}$ sean números reales positivos.

Demostración

Sea $\left(\left(\delta_1^{(m)}, \cdots, \delta_n^{(m)}\right)\right)_{m\in\mathbb{N}}$ una sucesión que converge al vector $\overline{0} \in \mathbb{R}^n$ y tal que $\delta_1^{(m)}, \cdots, \delta_n^{(m)}$ son números reales positivos. Entonces, para cualquier $k \in \{1, 2, \ldots, n\}$, se tiene:

$$\lim_{m \rightsquigarrow 0} \delta_k^{(m)} = 0.$$

Así que, para cada $k \in \{1, 2, \ldots, n\}$, existe una subsucesión $\left(\delta_k^{(m_j)}\right)_{j\in\mathbb{N}}$, de la sucesión $\left(\delta_k^{(m)}\right)_{m\in\mathbb{N}}$, la cual es decreciente.

Por lo tanto:

$$\lim_{m \rightsquigarrow \infty} F_{X_1,\ldots,X_n}\left(x_1 + \delta_1^{(m)}, \cdots, x_n + \delta_n^{(m)}\right)$$

$$= \lim_{j \rightsquigarrow \infty} F_{X_1,\ldots,X_n}\left(x_1 + \delta_1^{(m_j)}, \cdots, x_n + \delta_n^{(m_j)}\right)$$

$$= \lim_{j \rightsquigarrow \infty} P\left(\left[X_1 \leq x_1 + \delta_1^{(m_j)}, \ldots, X_n \leq x_n + \delta_n^{(m_j)}\right]\right)$$

$$= P\left(\bigcap_{j=1}^{\infty} \left[X_1 \leq x_1 + \delta_1^{(m_j)}, \ldots, X_n \leq x_n + \delta_n^{(m_j)}\right]\right)$$

$$= P([X_1 \leq x_1, \ldots, X_n \leq x_n])$$

$$= F_{X_1,\ldots,X_n}(x_1, \cdots, x_n).$$

∎

TEOREMA 13.10. *n variables aleatorias reales, $X_1, \ldots, X_n$, son independientes si y sólo si:*

$$F_{X_1,\ldots,X_n}(x_1, \cdots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

para cualquier vector $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

Demostración

Si $X_1, \ldots, X_n$, son independientes , la relación

$F_{X_1,\dots,X_n}(x_1,\cdots,x_n) = F_{X_1}(x_1)\, F_{X_2}(x_2)\cdots F_{X_n}(x_n)$

para cualquier vector $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, se sigue inmediatamente de la definición.

Inversamente, supongamos que $F_{X_1,\dots,X_n}(x_1,\cdots,x_n) = F_{X_1}(x_1)\, F_{X_2}(x_2)\cdots F_{X_n}(x_n)$, para cualquier vector $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

Definamos:

$$\mathcal{H}_1 = \left\{ A \subset \mathbb{R} : \begin{array}{l} A \text{ es boreliano y} \\ P[X_1 \in A, X_2 \leq x_2, \dots, X_n \leq x_n] \\ = P[X_1 \in A]\, P[X_2 \leq x_2] \cdots P[X_n \leq x_n] \\ \text{para cualquier vector } (x_2, \dots, x_n) \in \mathbb{R}^{n-1} \end{array} \right\}$$

$\mathcal{H}_1$ es entonces un d-sistema que contiene a todos los intervalos de la forma $(-\infty, x]$, los cuales forman un π-sistema que genera a los borelianos. Con base en el teorema de clases monótonas se concluye entonces que $\mathcal{H}_1$ contiene a todos los borelianos. Es decir:

$$P[X_1 \in A, X_2 \leq x_2, \dots, X_n \leq x_n] = P[X_1 \in A]\, P[X_2 \leq x_2] \cdots P[X_n \leq x_n]$$

para cualquier boreliano $A \subset \mathbb{R}$ y cualquier vector $(x_2, \dots, x_n) \in \mathbb{R}^{n-1}$.

Sea $j \in \{1, 2, \dots, n\}$ y supongamos que $A_1, A_2, \dots, A_j$ son j conjuntos borelianos de $\mathbb{R}$ tales que:

$$P[X_1 \in A_1, X_2 \in A_2, \dots, X_j \in A_j, X_{j+1} \leq x_{j+1}, X_{j+2} \leq x_{j+2}, \dots, X_n \leq x_n]$$

$$= P[X_1 \in A_1]\, P[X_2 \in A_2] \cdots P[X_j \in A_j]\, P[X_{j+1} \leq x_{j+1}]\, P[X_{j+2} \leq x_{j+2}] \cdots P[X_n \leq x_n]$$

para cualquier vector $(x_{j+1}, x_{j+2}, \dots, x_n) \in \mathbb{R}^{n-j}$.

Definamos:

$$\mathcal{H} = \left\{ B \subset \mathbb{R} : \begin{array}{l} B \text{ es boreliano y} \\ P[X_1 \in A_1, X_2 \in A_2, \dots, X_j \in A_j, X_{j+1} \in B, X_{j+2} \leq x_{j+2}, \dots, X_n \leq x_n] \\ = P[X_1 \in A_1]\, P[X_2 \in A_2] \cdots \\ P[X_j \in A_j]\, P[X_{j+1} \in B]\, P[X_{j+2} \leq x_{j+2}] \cdots P[X_n \leq x_n] \\ \text{para cualquier vector } (x_{j+2}, \dots, x_n) \in \mathbb{R}^{n-(j+1)} \end{array} \right\}$$

$\mathcal{H}$ es entonces un d-sistema que contiene a todos los intervalos de la forma $(-\infty, y]$, los cuales forman un π-sistema que genera a los borelianos. Con base en el teorema de clases monótonas, se concluye entonces que $\mathcal{H}$ contiene a todos los borelianos. Es decir:

$$P[X_1 \in A_1, X_2 \in A_2, \dots, X_j \in A_j, X_{j+1} \in B, X_{j+2} \leq x_{j+2}, \dots, X_n \leq x_n]$$

$$= P[X_1 \in A_1]\, P[X_2 \in A_2] \cdots P[X_j \in A_j]\, P[X_{j+1} \in B]\, P[X_{j+2} \leq x_{j+2}] \cdots P[X_n \leq x_n]$$

para cualquier boreliano $B \subset \mathbb{R}$ y cualquier vector $(x_{j+2}, \dots, x_n) \in \mathbb{R}^{n-(j+1)}$.

Partiendo de que $\mathcal{H}_1$ contiene a todos los borelianos de $\mathbb{R}$ y aplicando el resultado anterior $n-$ veces, se obtiene que:

$$P[X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n] = P[X_1 \in A_1] P[X_2 \in A_2] \cdots P[X_n \in A_n]$$

para cualquier colección $A_1, A_2, \ldots, A_n$ de n conjuntos borelianos de $\mathbb{R}$.

Por lo tanto, $X_1, \ldots, X_n$, son independientes.

■

CAPÍTULO 14

Esperanza y leyes de los grandes números

14.1. Esperanza de una variable aleatoria

Uno de los conceptos básicos de la teoría de la probabilidad es el de esperanza de una variable aleatoria. Su importancia es comparable con la del concepto mismo de probabilidad de un evento. De hecho, el concepto de esperanza y el de probabilidad surgieron en forma paralela cuando, a mediados del siglo $XVII$, se inicia el Cálculo de Probabilidades. Fue Christiaan Huygens quien introdujo este concepto en su libro *Du calcul dans les jeux de hasard*, publicado en el año 1657 ([**49**]). En esa época, Blaise Pascal y Pierre de Fermat habían resuelto algunos problemas de probabilidad, con métodos que sentarían las bases para el desarrollo de una nueva disciplina matemática, el Cálculo de Probabilidades. Huygens resolvió, con sus propios métodos, los problemas que antes habían resuelto Pascal y Fermat y algunos otros más.

Uno de los aspectos interesantes de la metodología utilizada por Huygens es que en ningún momento se consideran ahí probabilidades de eventos, todas las soluciones están basadas en el cálculo de esperanzas, lo cual hacía ver ya que este concepto podía tomarse como primario, previo incluso al de probabilidad de un evento, y a partir de él desarrollar la nueva disciplina. La historia no fue de ese modo pues el concepto que prevaleció como primario fue el de probabilidad. Sin embargo, la historia misma mostraría más adelante que esta dualidad de importancia, entre el concepto de esperanza y el de probabilidad, que se dio al inicio del desarrollo de la teoría de la probabilidad como disciplina matemática, tenía fuertes raíces pues al evolucionar el concepto de probabilidad, hasta fusionarse con el de medida en los primeros años de este siglo, resultó palpable la estrecha relación entre ambos conceptos, de tal manera que, efectivamente, cualquiera de los dos puede tomarse como punto de partida, quedando inmersos uno dentro del otro. Esto último ya no únicamente dentro del contexto de la teoría de la probabilidad, sino dentro del contexto más amplio de la teoría de la medida, donde el concepto de probabilidad corresponde al de medida y el de esperanza al de integral.

Definición 14.1. *Si X es una variable aleatoria no negativa, definimos la esperanza de X, $E[X]$, de la siguiente manera:*

$E[X] = \int_{\not\leq} X dP.$

DEFINICIÓN 14.2. *Diremos que la variable aleatoria X tiene esperanza finita si $E[|X|] < \infty$. En ese caso definimos su esperanza de la siguiente manera:*

$$E[X] = E[X^+] - E[X^-].$$

La Esperanza de una variable aleatoria tiene las mismas propiedades que la integral, es decir, se tiene la linealidad, el teorema de la convergencia monótona, el lema de Fatou, etcétera. Además, algunas propiedades útiles son las siguientes:

TEOREMA 14.1. *Si X es una variable aleatoria real no negativa, entonces $E[f(X)] = \int_0^\infty f(x)\, dF_X(x)$ para cualquier función boreliana $f : \overline{\mathbb{R}}^+ \mapsto \overline{\mathbb{R}}^+$.*

Demostración

X es una función definida sobre Ω con valores en $\overline{\mathbb{R}}^+$, así que, de acuerdo con la proposición 8.14, genera una medida μ_X sobre $\left(\overline{\mathbb{R}}^+, \beta\left(\overline{\mathbb{R}}^+\right)\right)$, la cual está definida por $\mu_X(B) = P[X \in B]$; además, si $f : \overline{\mathbb{R}}^+ \mapsto \overline{\mathbb{R}}^+$ es una función medible, se tiene:

$$\int_{\overline{\mathbb{R}}^+} f d\mu_X = \int_\Omega f \circ X dP.$$

Pero μ_X es la medida generada por la función de distribución de X, así que se tiene:

$$\int_{\overline{\mathbb{R}}^+} f dF_X = \int_\Omega f \circ X dP.$$

■

PROPOSICIÓN 14.1. *Sea X una variable aleatoria real no negativa, entonces:*

$$E[X] = \int_0^\infty [1 - F_X(x)]\, dx.$$

Demostración

$$E[X] = E\left[\int_0^X dx\right] = \int_\Omega \int_0^\infty I_{[0,X]}(x)\, dx dP = \int_0^\infty \int_\Omega I_{[0,X]}(x)\, dP dx$$

$$= \int_0^\infty P[X \geq x]\, dx = \int_0^\infty P[X > x]\, dx = \int_0^\infty [1 - F_X(x)]\, dx.$$

■

COROLARIO 14.1. $E[X^+] = \int_0^\infty P[X > x]\, dx$ *y* $E[X^-] = \int_{-\infty}^0 P[X < x]\, dx$.

Demostración

$$E[X^+] = \int_0^\infty P[X^+ > x]\, dx = \int_0^\infty P[X > x]\, dx.$$

$$E[X^-] = \int_0^\infty P[X^- > x]\, dx = \int_0^\infty P[X < -x]\, dx = \int_{-\infty}^0 P[X < x]\, dx.$$

■

COROLARIO 14.2. *Una variable aleatoria real X tiene esperanza finita si y sólo si las integrales $\int_0^\infty P[X > x]\, dx$ y $\int_{-\infty}^0 P[X < x]\, dx$ convergen.*

Corolario 14.3. *Si X es una variable aleatoria real de esperanza finita, entonces:*

$E[X] = \int_0^\infty P[X > x]\,dx - \int_{-\infty}^0 P[X < x]\,dx.$

Obsérvese que, como $F_X(x-) = P[X < x]$ y el conjunto de discontinuidades de F_X es a lo más infinbito numerable, $P[X < x] = F_X(x)$ excepto a lo más en un conjunto numerable, así que $\int_{-\infty}^0 P[X < x]\,dx = \int_{-\infty}^0 F_X(x)dx$. Además, $P[X > x] = 1 - F_X(x)$, así que entonces se tiene el siguiente resultado:

Teorema 14.2. *Sea X una variable aleatoria real, entonces X tiene esperanza finita si y sólo si $\int_0^\infty [1 - F_X(x)]\,dx < \infty$ y $\int_{-\infty}^0 F_X(x)dx < \infty$ y, en ese caso, se tiene:*

$E[X] = \int_0^\infty [1 - F_X(x)]\,dx - \int_{-\infty}^0 F_X(x)dx.$

Cuando se tiene $\int_0^\infty P[X > x]\,dx = \infty$ y $\int_{-\infty}^0 P[X < x]\,dx < \infty$, se define $E[X] = \infty$, mientras que cuando $\int_0^\infty P[X > x]\,dx < \infty$ y $\int_{-\infty}^0 P[X < x]\,dx = \infty$, se define $E[X] = -\infty$. Cuando ambas integrales sean divergentes, entonces la esperanza de X no está definida.

Proposición 14.2. *Sea X una variable aleatoria real no negativa y $\phi : \mathbb{R}^+ \to \mathbb{R}$ una función no decreciente, continua por la derecha y nula en $x = 0$, entonces:*

$E[\phi(X)] = \int_0^\infty P[X \geq x]\,d\phi(x).$

Demostración

$E[\phi(X)] = \int_\Omega \phi(X(\omega))\,dP(\omega) = \int_\Omega \left(\int_0^\infty I_{[0,X(\omega)]}(x)\,d\phi(x)\right) dP$

$= \int_0^\infty \left(\int_\Omega I_{[0,X(\omega)]}(x)\,dP\right) d\phi(x) = \int_0^\infty P[X \geq x]\,d\phi(x).$

■

Teorema 14.3. *Sean X y Y dos variables aleatorias independientes de esperanza finita, entonces XY también tiene esperanza finita y $E[XY] = E[X]\,E[Y]$.*

Demostración

$\varphi = \sum_{k=1}^m b_k I_{E_k}$, en donde $b_1, \ldots, b_m$ son números reales y $E_1, \ldots, E_m$ son conjuntos medibles.

Supongamos primero que X y Y son variables aleatorias discretas no negativas.

Sea V_X el conjunto de posibles valores de X, entonces:

$E[XY] = \int_0^\infty P[XY > z]\,dz$

$= \int_0^\infty \sum_{x \in V_X} P[XY > z, X = x]\,dz$

$= \sum_{\{x \in V_X : x > 0\}} \int_0^\infty P\left[Y > \frac{z}{x}, X = x\right] dz$

$= \sum_{\{x \in V_X : x > 0\}} \int_0^\infty xP[Y > y, X = x]\,dy$

$= \sum_{\{x \in V_X : x > 0\}} \int_0^\infty xP[X = x] \, P[Y > y] \, dy$

$= \sum_{\{x \in V_X : x > 0\}} xP[X = x] \int_0^\infty P[Y > y] \, dy$

$= E[X] \, E[Y]$.

Sean ahora X y Y dos variables aleatorias independientes no negativas. De acuerdo con el teorema 7.3, existen dos sucesiones $(X_n)_{n \in \mathbb{N}}$ y $(Y_n)_{n \in \mathbb{N}}$ de variables aleatorias discretas no negativas tales que $X_n \leq X_{n+1}$ y $Y_n \leq Y_{n+1}$ para cualquier $n \in \mathbb{N}$ y $\text{lím}_{n \rightsquigarrow \infty} X_n(\omega) = X(\omega)$ y $\text{lím}_{n \rightsquigarrow \infty} Y_n(\omega) = Y(\omega)$ para cualquier $\omega \in \Omega$, lo cual implica $X_n Y_n \leq X_{n+1} Y_{n+1}$ para cualquier $n \in \mathbb{N}$ y $\text{lím}_{n \rightsquigarrow \infty} (X_n Y_n)(\omega) = (XY)(\omega)$ para cualquier $\omega \in \Omega$. Además, de acuerdo con la demostración del teorema, para cada $n \in \mathbb{N}$, existe una función medible $f_n : \mathbb{R} \mapsto \mathbb{R}$ tal que $X_n = f_n(X)$ y $Y_n = f_n(Y)$, de manera que, por la proposición 13.4, X_n y Y_n son independientes.

De manera que, por el teorema de la convergencia monótona, se tiene:

$E[XY] = \text{lím}_{n \rightsquigarrow \infty} E[X_n Y_n]$

$= \text{lím}_{n \rightsquigarrow \infty} E[X_n] \, \text{lím}_{n \rightsquigarrow \infty} E[Y_n] = E[X] \, E[Y]$.

Finalmente, si X y Y dos variables aleatorias independientes cualesquiera de esperanza finita, se tiene:

$|XY| = |(X^+ - X^-)(Y^+ - Y^-)| \leq (X^+ + X^-)(Y^+ + Y^-)$.

Pero, siendo $X^+ + X^-$ y $Y^+ + Y^-$ variables aleatorias no negativas de esperanza finita, su producto también lo es, de manera que XY tiene esperanza finita.

Además, por la proposición 13.4, $X^+ = \text{máx}(X, 0)$ y $Y^+ = \text{máx}(Y, 0)$ son independientes y, de la misma manera, lo son X^+ y Y^-, X^- y Y^+ y X^- y Y^-. Por lo tanto:

$E[XY] = E[(X^+ - X^-)(Y^+ - Y^-)] = E[X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+]$

$= E[X^+Y^+] + E[X^-Y^-] - E[X^+Y^-] - E[X^-Y^+]$

$= E[X^+] \, E[Y^+] + E[X^-] \, E[Y^-] - E[X^+] \, E[Y^-] - E[X^-] \, E[Y^+]$

$= (E[X^+] - E[X^-])(E[Y^+] - E[Y^-]) = E[X] \, E[Y]$.

■

Un razonamiento de inducción permite demostrar el siguiente corolario:

Corolario 14.4. *Sean $X_1, \ldots, X_n$ n variables aleatorias independientes de esperanza finita, entonces $\prod_{k=1}^n X_k$ también tiene esperanza finita y $E[\prod_{k=1}^n X_k] = \prod_{k=1}^n E[X_k]$.*

14.2. Varianza y covarianza

DEFINICIÓN 14.3 (**Varianza**). *Sea X una variable aleatoria de esperanza finita. Se define la varianza de X, $Var(X)$, mediante la relación:*

$Var(X) = E\left[(X - E(X))^2\right]$.

A la raíz cuadrada no negativa de la varianza se le llama la ***desviación estándar*** *de X.*

La varianza de una variable aleatoria mide entonces el alejamiento de los valores de X de su esperanza. También se acostumbra decir que la varianza es una medida de la dispersión de los valores de la variable aleatoria.

DEFINICIÓN 14.4 (**Varianza finita**). *Diremos que una variable aleatoria X tiene varianza finita si se cumplen las siguientes dos condiciones:*

(i) *X tiene esperanza finita.*
(ii) *$(X - E[X])^2$ tiene esperanza finita.*

PROPOSICIÓN 14.3. *Una variable aleatoria X tiene varianza finita si y sólo si X^2 tiene esperanza finita.*

Demostración

Se tiene $X^2 = (X - E[X])^2 + 2XE[X] - (E[X])^2$, así que si X tiene varianza finita, entonces X^2 tiene esperanza finita.

Supongamos ahora que X^2 tiene esperanza finita.

Se tiene $|X| \leq 1 + X^2$ y $(X - E[X])^2 = X^2 - 2XE[X] - (E[X])^2$. De manera que tanto X como $(X - E[X])^2$ tienen esperanza finita. Es decir, X tiene varianza finita.

PROPOSICIÓN 14.4. *Sea X una variable aleatoria de esperanza finita, entonces:*

$Var(X) = E[X^2] - (E[X])^2$.

Demostración

Si X no tiene varianza finita entonces X^2 no tiene esperanza finita, así que se cumple la igualdad. Si X tiene varianza finita, entonces X^2 tiene esperanza finita y se tiene:

$Var(X) = E\left[(X - E(X))^2\right] = E\left[X^2 - 2XE(X) + (E[X])^2\right]$

$= E[X^2] - 2(E[X])^2 + (E[X])^2$

$= E[X^2] - (E[X])^2$.

■

PROPOSICIÓN 14.5. *Sean X y Y dos variables aleatorias de varianza finita. Entonces, XY tiene esperanza finita.*

Demostración

Para cualquier par de números reales x y y, se tiene $|xy| \leq \frac{1}{2}(x^2 + y^2)$. Así que, $|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2$.

Por lo tanto, XY tiene esperanza finita. ■

COROLARIO 14.5. *Si X y Y son dos variables aleatorias de varianza finita y a y b son dos números reales cualesquiera, entonces $aX + bY$ tiene varianza finita.*

Demostración

$(aX + bY)^2 = a^2X^2 + 2abXY + b^2Y^2$, así que, por las proposiciones 14.5 y 14.3, $aX + bY$ tiene varianza finita.

■

DEFINICIÓN 14.5 (**Covarianza**). *Sean X y Y dos variables aleatorias de varianza finita. Se define la covarianza de X y Y, $Cov(X, Y)$, mediante la relación:*

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

PROPOSICIÓN 14.6. *Sean X y Y dos variables aleatorias independientes de varianza finita, entonces $Cov(X, Y) = 0$.*

Demostración

El resultado es inmediato pues X y Y son independientes y tienen esperanza finita, así que $E[XY] = E[X]E[Y]$.

■

El siguiente ejemplo muestra que la covarianza entre dos variables aleatorias puede ser cero sin que éstas sean independientes.

EJEMPLO 14.1. *Sea X una variable aleatoria con función de densidad dada por:*

$$f(x) = \begin{cases} \frac{1}{4} & si\ x \in \{-1, 1\} \\ \frac{1}{2} & si\ x = 0 \\ 0 & en\ otro\ caso \end{cases}$$

y sea Z una variable aleatoria, independiente de X, con distribución uniforme en el conjunto $\{-1, 1\}$. Definamos la variable aleatoria Y de la siguiente manera:

$$Y = \begin{cases} Z & si\ X = 0 \\ 0 & en\ otro\ caso \end{cases}$$

Se tiene entonces:

$f_Y(y) = P[Y = y] = P[Y = y, X = 0] + P[Y = y, X \neq 0]$

$= P[Z = y, X = 0] + P[Y = y, X \neq 0]$

$= P[Z = y] P[X = 0] + I_{\{0\}}(y) P[X \neq 0]$

$= \frac{1}{4} I_{\{-1,1\}}(y) + \frac{1}{2} I_{\{0\}}(y) = \begin{cases} \frac{1}{4} & si\ y \in \{-1, 1\} \\ \frac{1}{2} & si\ y = 0 \\ 0 & en\ otro\ caso \end{cases}$

$f_{X,Y}(x, y) = P[X = x, Y = y]$

$= P[X = x, Y = y, X = 0] + P[X = x, Y = y, X \neq 0]$

$= I_{\{0\}}(x) P[Z = y, X = 0] + I_{\{0\}}(y) I_{\{-1,1\}}(x) P[X = x]$

$= I_{\{0\}}(x) P[Z = y] P[X = 0] + I_{\{0\}}(y) I_{\{-1,1\}}(x) P[X = x]$

$= \frac{1}{4} I_{\{0\}}(x) I_{\{-1,1\}}(y) + \frac{1}{4} I_{\{0\}}(y) I_{\{-1,1\}}(x)$

$= \begin{cases} \frac{1}{4} & si\ x = 0, y \in \{-1, 1\} \ \ ó\ y = 0, x \in \{-1, 1\} \\ 0 & en\ otro\ caso \end{cases}$

Así que, por ejemplo, $P[X = 1, Y = 1] \neq P[X = 1] P[Y = 1]$*, de manera que* X *y* Y *no son independientes. Por otro lado,* $E[X] = E[Y] = E[XY] = 0$*, de manera que* $Cov(X, Y) = 0$.

PROPOSICIÓN 14.7. *Sean* X *y* Y *dos variables aleatorias de varianza finita y* a *y* b *dos números reales cualesquiera. Entonces* $Cov(aX, bY) = abCov(X, Y)$.

Demostración

$Cov(aX, bY) = E[aXbY] - E[aX] E[bY] = ab(E[XY] - E[X] E[Y])$

$= abCov(X, Y)$.

■

TEOREMA 14.4. *Sean* $X, X_1, \ldots, X_n$ $n+1$ *variables aleatorias de esperanza finita. Entonces:*

(i) $Var(X) = 0$ *si y sólo si existe una constante* c *tal que* $P[X = c] = 1$.
(ii) $Var(aX + b) = a^2 Var(X)$ *para cualesquiera constantes* a *y* b.
(iii) *Si* $X_1, \ldots, X_n$ *tienen varianza finita, entonces* $\sum_{i=1}^n X_i$ *también tiene varianza finita y:*

$$Var(\textstyle\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + 2 \sum_{\{i,j \in \{1,\ldots,n\}: i<j\}} Cov(X_i, Y_j).$$

Demostración

1. $Var(X) = 0$ si y sólo si $E\left[(X - E(X))^2\right] = 0$, lo cual, por la proposición 7.13, ocurre si y sólo si $P\left[(X - E(X))^2 = 0\right] = 1$, es decir, $P[X = E(X)] = 1$.

2. $Var(aX + b) = E\left[(aX - aE[X])^2\right] = a^2 E\left[(X - E[X])^2\right] = a^2 Var(X)$.

3. Que $\sum_{i=1}^{n} X_i$ tiene varianza finita se sigue del corolario 14.5 y un razonamiento de inducción. Además:

$Var(\sum_{i=1}^{n} X_i) = E\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} E[X_i]\right)^2\right] = E\left[\left(\sum_{i=1}^{n} (X_i - E[X_i])\right)^2\right]$

$= \sum_{i=1}^{n} E\left[(X_i - E[X_i])^2\right] + 2\sum_{\{i,j\in\{1,\ldots,n\}:i<j\}} E\left[(X_i - E[X_i])(X_j - E[X_j])\right]$

$= \sum_{i=1}^{n} Var(X_i) + 2\sum_{\{i,j\in\{1,\ldots,n\}:i<j\}} Cov(X_i, Y_j)$. ∎

COROLARIO 14.6. *Sean $X_1, \ldots, X_n$ n variables aleatorias independientes y de varianza finita, entonces $\sum_{i=1}^{n} X_i$ también tiene varianza finita y $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$.*

TEOREMA 14.5 (**Desigualdad de Cauchy-Schwarz**). *Sean X y Y dos variables aleatorias cualesquiera, entonces:*

$E[|XY|] \leq \sqrt{E[X^2]}\sqrt{E[Y^2]}$.

Además, si X y Y tienen varianza finita, entonces $|E[XY]| = \sqrt{E[X^2]}\sqrt{E[Y^2]}$ si y sólo si existen constantes a y b tales que por lo menos una de ellas es distinta de cero y $P[aX + bY = 0] = 1$.

Demostración

Si $E[X^2] = \infty$ o $E[Y^2] = \infty$ la desigualdad es obvia.

Supongamos ahora que $E[X^2] < \infty$ y $E[Y^2] < \infty$, es decir, que tanto X como Y tienen varianza finita.

Sea $\alpha = (E[Y^2])^{\frac{1}{2}}$ y $\beta = (E[X^2])^{\frac{1}{2}}$.

Si $\alpha = 0$, se tiene $E[X^2] = 0$, de manera que:

$P[|XY| = 0] \geq P[X = 0] = P[X^2 = 0] = 1$

Por lo tanto, $E[|XY|] = 0$. Así que se cumple la desigualdad.

De la misma manera, si $\beta = 0$, entonces $E[|XY|] = 0$. Así que se cumple la desigualdad.

Supongamos ahora que $\alpha > 0$ y $\beta > 0$.

Sabemos que $\alpha|X| - \beta|Y|$ tiene varianza finita y se tiene:

$0 \leq E\left[(\alpha|X| - \beta|Y|)^2\right] = \alpha^2 E[X^2] + \beta^2 E[Y^2] - 2\alpha\beta E[|XY|] = 2\alpha^2\beta^2 - 2\alpha\beta E[|XY|]$.

Así que, $\alpha\beta - E[|XY|] \geq 0$. Es decir, $E[|XY|] \leq \alpha\beta$.

Para la segunda parte, supongamos primero que X y Y tienen varianza finita y que $|E[XY]| = \sqrt{E[X^2]}\sqrt{E[Y^2]}$.

Definiendo, como antes, $\alpha = (E[Y^2])^{\frac{1}{2}}$ y $\beta = (E[X^2])^{\frac{1}{2}}$, se tiene:

Si $\alpha = 0$ y $\beta = 0$, entonces $P[X = 0] = P[Y = 0] = 1$. Por lo tanto $P[X = 0, Y = 0] = 1$. De manera que, tomando en consideración que $P[X = 0, Y = 0] \leq P[X + Y = 0]$, se tiene $P[X + Y = 0] = 1$. Es decir, se tiene el resultado deseado con $a = b = 1$.

Si $\alpha \neq 0$ ó $\beta \neq 0$ se tienen los siguientes dos casos:

Si $E[XY] > 0$, entonces:

$0 \leq E\left[(\alpha X - \beta Y)^2\right] = 2\alpha^2\beta^2 - 2\alpha\beta E[XY] = 0.$

Así que, $E\left[(\alpha X - \beta Y)^2\right] = 0$, de lo cual se sigue $P[\alpha X - \beta Y = 0] = 1$.

Es decir, se tiene el resultado deseado con $a = \alpha$ y $b = -\beta$.

Si $E[XY] < 0$, entonces:

$0 \leq E\left[(\alpha X + \beta Y)^2\right] = 2\alpha^2\beta^2 + 2\alpha\beta E[XY] = 0.$

Así que, $E\left[(\alpha X + \beta Y)^2\right] = 0$, de lo cual se sigue $P[\alpha X + \beta Y = 0] = 1$.

Es decir, se tiene el resultado deseado con $a = \alpha$ y $b = \beta$.

Finalmente, supongamos que existen constantes a y b tales que por lo menos una de ellas es distinta de cero y $P[aX + bY = 0] = 1$. Supongamos, por ejemplo, que $a \neq 0$, entonces $P\left[X = -\frac{b}{a}Y\right] = 1$. Así que:

$(E[XY])^2 = \frac{b^2}{a^2}(E[Y^2])^2 = E\left[\left(-\frac{b}{a}Y\right)^2\right] E[Y^2] = E[X^2]\,E[Y^2].$

∎

Corolario 14.7. *Sean X y Y dos variables aleatorias de varianza finita. Entonces:*

$|Cov(X, Y)| \leq \sqrt{Var(X)}\sqrt{Var(Y)}.$

Además, la igualdad se cumple si y sólo si existen constantes a, b y c tales que a y b no son ambas cero y $P[aX + bY = c] = 1$.

Demostración

Utilizando la proposición 14.5, se tiene:

$|Cov(X, Y)| = |E[(X - E[X])(Y - E[Y])]| \leq E[|X - E[X]|\,|Y - E[Y]|]$

$\leq \sqrt{E\left[(X - E[X])^2\right]}\sqrt{E\left[(Y - E[Y])^2\right]} = \sqrt{Var(X)}\sqrt{Var(Y)}.$

Si la igualdad se cumple, entonces se tiene

$$|E[(X - E[X])(Y - E[Y])]| = \sqrt{E\left[(X - E[X])^2\right]}\sqrt{E\left[(Y - E[Y])^2\right]}.$$

De manera que, nuevamente por la proposición 14.5, existen constantes a y b tales que no son ambas cero y $P[a(X - E[X]) + b(Y - E[Y]) = 0] = 1$. Es decir:

$$P[aX + bY = c] = 1,$$

donde $c = aE[X] + bE[Y]$.

Supongamos ahora que existen constantes a, b y c tales que a y b no son ambas cero y

$P[aX + bY = c] = 1$. Entonces $E[aX + bY - c] = 0$, de lo cual se sigue $c = E[aX + bY]$. De manera que se tiene:

$$P[a(X - E[X]) + b(Y - E[Y]) = 0] = 1.$$

Así que, por la proposición 14.5, se tiene:

$$|Cov(X,Y)| = |E[(X - E[X])(Y - E[Y])]| = \sqrt{Var(X)}\sqrt{Var(Y)}.$$

■

14.3. Desigualdad de Chebyshev

El gran impulso para el desarrollo de una teoría de la probabilidad, que le haría ganar un lugar dentro de las matemáticas, proviene de los llamados teoremas límite, los cuales se refieren al comportamiento a largo plazo de sucesiones de variables aleatorias. El primero de estos resultados, que para algunos autores marca verdaderamente el inicio de la historia de la teoría de la probabilidad, se debe a Jacques Bernoulli, quien dedicó 20 años de su vida a la búsqueda de una prueba matemática de la relación que existe entre la probabilidad de un evento y la frecuencia relativa con la que éste ocurre en una serie grande de repeticiones del correspondiente experimento aleatorio. El resultado, conocido como teorema de Bernoulli, se publicó en el año 1713, ocho años después de la muerte de su autor.

Puede decirse que, a partir de la publicación del teorema de Bernoulli, el motor de desarrollo de la teoría de la probabilidad fue la búsqueda de resultados que permitieran mejorar y generalizar ese teorema. Vendrían después los teoremas de de Moivre y de Poisson, relativos a la aproximación de una distribución binomial mediante una distribución normal y una distribución Poisson, respectivamente, los cuales fueron publicados en los años 1730 y 1800, respectivamente.

Este proceso continuaría desarrollándose y recibiría un gran impulso, entre 1870 y 1900, con los trabajos de la llamada escuela rusa, representada por Pafnuty Lvovich Chebyshev ([**18**], [**19**], [**20**]), Andrei Andreyevich Markov ([**66**], [**67**], [**68**], [**69**]) y Aleksandr Mikhailovich Lyapunov ([**63**], [**64**]), entre otros, los cuales conducirían a la forma general que se dio a los teoremas límite, entre 1900 y 1930, con la formulación de las leyes de los grandes números y el teorema central del límite, tanto en su forma clásica, relativa a la convergencia a la distribución normal, como en su forma moderna, relativa a la convergencia a cualquier otro tipo de distribución, sobresaliendo en este periodo los trabajos de Aleksandr Yakovlevich Khintchine, Andrey Nikolaevich Kolmogorov, J. W. Lindeberg, William Feller y Paul Pierre Lévy, entre otros.

Como puede verse, fueron más de 200 años de historia de la teoría de la probabilidad guiada por el estudio de los teoremas límite.

PROPOSICIÓN 14.8. *Sea X cualquier variable aleatoria y ε cualquier número real positivo, entonces:*

$P\left[|X| \geq \varepsilon\right] \leq \frac{1}{\varepsilon} E\left[|X|\right]$.

Demostración

$E\left[|X|\right] = \int_0^\infty \left[1 - F_{|X|}(x)\right] dx = \int_0^\varepsilon \left[1 - F_{|X|}(x)\right] dx + \int_\varepsilon^\infty \left[1 - F_{|X|}(x)\right] dx$

$\geq \int_0^\varepsilon \left[1 - F_{|X|}(x)\right] dx = \int_0^\varepsilon P\left[|X| > x\right] dx \geq \int_0^\varepsilon P\left[|X| \geq \varepsilon\right] dx = \varepsilon P\left[|X| \geq \varepsilon\right]$.

■

COROLARIO 14.8. *Sea X cualquier variable aleatoria y ε cualquier número real positivo, entonces:*

$P\left[|X| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2} E\left[X^2\right]$.

Demostración

$P\left[|X| \geq \varepsilon\right] = P\left[X^2 \geq \varepsilon^2\right] \leq \frac{1}{\varepsilon^2} E\left[X^2\right]$.

■

COROLARIO 14.9 (**Desigualdad de Chebyshev**). *Sea X cualquier variable aleatoria de esperanza finita y ε cualquier número real positivo, entonces:*

$P\left[|X - E\left[X\right]| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2} Var\left[X\right]$.

14.4. Léy débil de los grandes números

TEOREMA 14.6 (**Ley débil de los grandes números de Chebyshev**). *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de varianza finita. Entonces:*

$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{P} \mu$,

donde μ es la esperanza común de $X_1, X_2, \ldots$.

Demostración

Para cada $n \in \mathbb{N}$, sea $Y_n = \frac{X_1+\cdots+X_n}{n}$. Entonces Y_n es una variable aleatoria de varianza finita y esperanza μ. De manera que, por la desigualdad de Chebyshev, se tiene:

$P\left[|Y_n - \mu| > \varepsilon\right] \leq \frac{1}{\varepsilon^2} Var\left[Y_n\right] = \frac{\sigma^2}{n\varepsilon^2}$,

donde σ^2 es la varianza común de $X_1, X_2, \ldots$. Tomando límites cuando $n \rightsquigarrow \infty$ se tiene entonces el resultado.

■

Lema 14.1. *Si $f : [0,\infty) \mapsto \mathbb{R}$ es una función decreciente y no negativa tal que $\int_0^\infty f(x)dx < \infty$ y (a_n) una sucesión creciente de números reales positivos tal que* $\text{lím}_{n\rightsquigarrow\infty} a_n = \infty$, *entonces* $\text{lím}_{n\rightsquigarrow\infty} a_n f(a_n) = 0$.

Demostración

La sucesión (s_n), en donde $s_n = \sum_{\{k\in\mathbb{N}:k\leq a_n\}} f(k)$, es no decreciente y se tiene:

$s_n = \sum_{\{k\in\mathbb{N}:k\leq a_n\}} \int_{k-1}^{k} f(k)dx \leq \sum_{\{k\in\mathbb{N}:k\leq a_n\}} \int_{k-1}^{k} f(x)dx$

$\leq \int_0^{a_n} f(x)dx \leq \int_0^\infty f(x)dx$.

Así que (s_n) converge y es, por lo tanto, una sucesión de Cauchy.

Entonces, dada $\varepsilon > 0$ existe un número natural M tal que si $n \geq m \geq M$ entonces $s_n - s_m < \frac{\varepsilon}{2}$, es decir $\sum_{\{k\in\mathbb{N}:a_m<k\leq a_n\}} f(k) < \frac{\varepsilon}{2}$.

Sea ahora N tal que $a_n > 2(a_M + 1)$ para cualquier $n > N$, se tiene entonces, para $n > N$, $a_n - 2(a_M + 1) > 0$ y $(a_n - a_M - 1)f(a_n) \leq \sum_{\{k\in\mathbb{N}:a_M<k\leq a_n\}} f(k) < \frac{\varepsilon}{2}$. Así que:

$a_n f(a_n) < 2(a_n - a_M - 1)f(a_n) < \varepsilon$,

lo cual prueba el resultado.

■

Proposición 14.9. *Si X es una variable aleatoria de esperanza finita y (a_n) una sucesión creciente de números reales positivos tal que* $\text{lím}_{n\rightsquigarrow\infty} a_n = \infty$, *entonces:*

$\text{lím}_{n\rightsquigarrow\infty} a_n P\left[X > a_n\right] = \text{lím}_{n\rightsquigarrow\infty} a_n P\left[X < -a_n\right] = 0.$

Demostración

Como X tiene esperanza finita, se tiene:

$\int_0^\infty P\left[X > x\right] dx = \int_0^\infty \left[1 - F_X(x)\right] dx < \infty$,

$\int_0^\infty P\left[X < -x\right] dx \leq \int_0^\infty P\left[X \leq -x\right] dx = \int_0^\infty F_X(-x) < \infty$.

Además, las funciones $x \mapsto P[X > x]$ y $x \mapsto P[X < -x]$ son no negativas y decrecientes en el intervalo $[0, \infty)$.

El resultado se sigue entonces del lema 14.1.

■

LEMA 14.2. *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ y (a_n) una sucesión creciente de números reales positivos tal que $\lim_{n \rightsquigarrow \infty} a_n = \infty$. Para $n, k \in \mathbb{N}$, definamos:*

$$Y_k^n = \begin{cases} X_k & si\ |X_k| \leq a_n \\ 0 & en\ otro\ caso \end{cases}.$$

Entonces, fijando n, las variables aleatorias $Y_1^n, Y_2^n, \ldots$ tienen la misma distribución. Además,si μ_n es la esperanza común de $Y_1^n, Y_2^n, \ldots$, entonces $\lim_{n \rightsquigarrow \infty} \mu_n = \mu$.

Demostración

$$F_{Y_k^n}(x) = P[Y_k^n \leq x] = P[Y_k^n \leq x, |X_k| \leq a_n] + P[Y_k^n \leq x, |X_k| > a_n]$$

$$= P[X_k \leq x, |X_k| \leq a_n] + P[Y_k^n \leq x, |X_k| > a_n]$$

$$= \begin{cases} 0 & \text{si } x < -a_n \\ P[-a_n \leq X_k \leq x] & \text{si } -a_n \leq x < 0 \\ P[-a_n \leq X_k \leq x] + P[|X_k| > a_n] & \text{si } 0 \leq x \leq a_n \\ 1 & \text{si } x > a_n \end{cases}$$

$$= \begin{cases} 0 & \text{si } x < -a_n \\ P[-a_n \leq X_k \leq x] & \text{si } -a_n \leq x < 0 \\ P[X_k \leq x] + P[X_k > a_n] & \text{si } 0 \leq x \leq a_n \\ 1 & \text{si } x > a_n \end{cases}$$

$$= \begin{cases} 0 & \text{si } x < -a_n \\ F_{X_k}(x) - P[X_k < -a_n] & \text{si } -a_n \leq x < 0 \\ F_{X_k}(x) + P[X_k > a_n] & \text{si } 0 \leq x \leq a_n \\ 1 & \text{si } x > a_n \end{cases}$$

De manera que, fijando n, las variables aleatorias $Y_1^n, Y_2^n, \ldots$ tienen la misma distribución. Además:

$$\mu_n = E[Y_1^n] = \int_0^\infty \left[1 - F_{Y_1^n}(x)\right] dx - \int_0^\infty F_{Y_1^n}(-x) dx$$

$$= \int_0^{a_n} \left[1 - F_{Y_1^n}(x)\right] dx - \int_0^{a_n} F_{Y_1^n}(-x) dx$$

$$= \int_0^{a_n} [1 - F_{X_1}(x) - P[X_1 > a_n]] dx - \int_0^{a_n} [F_{X_1}(-x) - P[X_1 < -a_n]] dx$$

$$= \int_0^{a_n} [1 - F_{X_1}(x)] dx - a_n P[X_1 > a_n] - \int_0^{a_n} F_{X_1}(-x) dx + a_n P[X_1 < -a_n]$$

$= \int_0^{a_n} [1 - F_{X_1}(x)]\, dx - \int_0^{a_n} F_{X_1}(-x) dx + a_n P[X_1 < -a_n] - a_n P[X_1 > a_n]$.

Así que, utilizando la proposición 14.9, $\lim_{n \rightsquigarrow \infty} \mu_n = E[X_1] = \mu$.

■

El siguiente resultado fue demostrado por Aleksandr Yakovlevich Khintchine en el año 1928 ([**51**]):

TEOREMA 14.7 (**Ley débil de los grandes números de Khintchine**). *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ. Entonces:*

$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mu$.

Demostración

Sea ν es el valor común de $E[|X_1|]$, $E[|X_2|], \ldots$. Si $\nu = 0$, el resultado es trivial. Supongamos entonces que $\nu > 0$.

Dada $\delta > 0$, definamos, para $n, k \in \mathbb{N}$:

$$a_n = \frac{\delta \varepsilon^2}{8\nu} n \text{ y } Y_k^n = \begin{cases} X_k & \text{si } |X_k| \leq a_n \\ 0 & \text{en otro caso} \end{cases}$$

Por el lema 14.2, fijando n, las variables aleatorias $Y_1^n, Y_2^n, \ldots$ tienen la misma distribución y si μ_n es la esperanza común de $Y_1^n, Y_2^n, \ldots$, entonces $\lim_{n \rightsquigarrow \infty} \mu_n = \mu$.

Por otra parte, para cualesquiera $n, k \in \mathbb{N}$, se tiene $(Y_k^n)^2 \leq a_n^2$, así que Y_k^n tiene varianza finita.

Además, $|Y_k^n| \leq |X_k|$ y $|Y_k^n| \leq a_n$, así que, si σ_n^2 es la varianza común de $Y_1^n, Y_2^n, \ldots$, se tiene:

$$\sigma_n^2 \leq E\left[(Y_k^n)^2\right] \leq E[a_n |X_k|] = a_n E[|X_k|] = \frac{\delta \varepsilon^2}{8\nu} n E[|X_k|] = \frac{\delta n \varepsilon^2}{8}.$$

Ahora bien, como $\lim_{n \rightsquigarrow \infty} \mu_n = \mu$ y $\lim_{n \rightsquigarrow \infty} a_n P[X_1 > a_n] = 0$, existe N tal que $|\mu_n - \mu| < \frac{\varepsilon}{2}$ y $a_n P[X_1 > a_n] < \frac{\delta^2}{2}$ para cualquier $n > N$.

Entonces, para $n > N$, se tiene:

$$P\left[\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right]$$

$$\leq P\left[\left|\frac{Y_1^n + \cdots + Y_n^n}{n} - \mu\right| > \varepsilon\right] + P[Y_k^n \neq X_k \text{ para alguna } k \leq n]$$

$$\leq P\left[\left|\frac{Y_1^n + \cdots + Y_n^n}{n} - \mu_n\right| > \frac{\varepsilon}{2}\right] + P[Y_k^n \neq X_k \text{ para alguna } k \leq n].$$

Pero, por la desigualdad de Chebyshev, se tiene:

$P\left[\left|\frac{Y_1^n+\cdots+Y_n^n}{n}-\mu_n\right|>\frac{\varepsilon}{2}\right]\leq\frac{4\sigma_n^2}{n\varepsilon^2}\leq\frac{\delta}{2}$.

Además:

$P\left[Y_k^n \neq X_k \text{ para alguna } k\leq n\right]\leq\sum_{k=1}^n P\left[Y_k^n\neq X_k\right]$

$=\sum_{k=1}^n P\left[|X_k|>a_n\right]=nP\left[X_1>a_n\right]$

$=\frac{n}{a_n}a_nP\left[X_1>a_n\right]=\frac{1}{\delta}a_nP\left[X_1>a_n\right]<\frac{\delta}{2}$.

Así que:

$P\left[\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|>\varepsilon\right]\leq\frac{\delta}{2}+\frac{\delta}{2}=\delta$,

lo cual prueba el resultado. ■

El método utilizado por Khintchine en la proposición anterior es conocido como el **método de truncación**. Fue introducido por Markov en el año 1913 con relación a un teorema de Aleksandr Mikhailovich Lyapunov, el cual generaliza el teorema de de Moivre.

14.5. Ley fuerte de los grandes números

Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de varianza finita y esperanza común μ. La ley débil de los grandes números establece que $\frac{X_1+\cdots+X_n}{n}\xrightarrow{P}\mu$. En el año 1930 Andrey Nikolaevich Kolmogorov mostró que este resultado puede mejorarse demostrando que la convergencia a μ se da no sólo en probabilidad sino también con probabilidad 1, la cual, como ya vimos, es un tipo de convergencia más fuerte.

Como vimos antes, la demostración de que la sucesión $Y_n=\frac{X_1+\cdots+X_n}{n}$ converge a μ en probabilidad está basada en la desigualdad de Chebyshev, de la cual se obtiene que $P\left[|Y_n-\mu|>\varepsilon\right]\leq\frac{K}{n}$, en donde K es una constante. De la proposición 10.3 puede verse, que para demostrar que la sucesión Y_n converge a μ con probabilidad 1 bastaría con demostrar que $\sum_{n=1}^\infty P\left[|Y_n-\mu|>\varepsilon\right]<\infty$ para cualquier $\varepsilon>0$. Para probar esto no basta con aplicar la desigualdad de Chebyshev puesto que ésta únicamente establece que $P\left[|Y_n-\mu|>\varepsilon\right]\leq\frac{K}{n}$ y la serie $\sum_{n=1}^\infty\frac{1}{n}$ no es convergente.

El resultado de Kolmogorov tiene su origen en el teorema de Borel, publicado en el año 1909, el cual se enuncia y demuestra a continuación (la demostración no es la original de Borel, sino la de Hausdorff):

TEOREMA 14.8 (Teorema de Borel). *Sea $\mathcal{E}$ un experimento aleatorio y A un evento relativo a ese experimento, de probabilidad igual a p. Consideremos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento $\mathcal{E}$, de tal manera que cada repetición es independiente de las otras. Sea X_n el número de veces que ocurre el evento A en las primeras n repeticiones del experimento, entonces $\frac{X_n}{n}\xrightarrow{c.s.}p$.*

Demostración

Sabemos que X_n tiene distribución binomial de parámetros n y p. Así que:

$E[X_n] = np,$

$E[X_n^2] = np + n(n-1)p^2,$

$E[X_n^3] = np + 3n(n-1)p^2 + n(n-1)(n-2)p^3,$

$E[X_n^4] = np + 7n(n-1)p^2 + 6n(n-1)(n-2)p^3 + n(n-1)(n-2)(n-3)p^4.$

Por lo tanto:

$$E\left[\left(\frac{X_n}{n} - p\right)^4\right] = \frac{E[X_n^4]}{n^4} - 4\frac{E[X_n^3]}{n^3}p + 6\frac{E[X_n^2]}{n^2}p^2 - 4\frac{E[X_n]}{n}p^3 + p^4$$

$$= \frac{1}{n^3}p(1-p)[3np(1-p) - 6p(1-p) + 1] < \frac{1}{4n^3}\left(\frac{3n}{4} + n\right) < \frac{1}{n^2}.$$

Sabemos además que si X es cualquier variable aleatoria y ε cualquier número real positivo, entonces $P[|X| \geq \varepsilon] \leq \frac{1}{\varepsilon}E[|X|]$, así que:

$$P\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right] \leq \frac{E\left[\left(\frac{X}{n} - p\right)^4\right]}{\varepsilon^4} < \frac{1}{n^2\varepsilon^4}.$$

La serie $\sum_{n=1}^{\infty} P\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right]$ es entonces convergente para cualquier $\varepsilon > 0$. Así que, por el corolario 10.3, $\frac{X_n}{n} - p \xrightarrow{c.s.} 0$, es decir, $\frac{X_n}{n} \xrightarrow{c.s.} p$.

■

El teorema de Borel equivale a decir que si $X_1, X_2, \ldots$ es una sucesión de variables aleatorias independientes, todas con distribución Bernoulli de parámetro p, entonces $\frac{X_1 + \cdots + X_n}{n} \xrightarrow{c.s.} p$.

El método de Kolmogorov para probar la convergencia con probabilidad 1 de la sucesión $Y_n = \frac{X_1 + \ldots + X_n}{n}$ está basado en una desigualdad más general que la de Chebyshev y que él mismo demuestra, por lo cual es llamada la desigualdad de Kolmogorov. Aquí daremos una versión ligeramente modificada de la demostración original.

TEOREMA 14.9 (**Desigualdad de Kolmogorov**). *Sean $X_1, \ldots, X_n$ n variables aleatorias independientes de varianza finita y ε cualquier número real positivo, entonces:*

$$P\left[\max_{1 \leq j \leq n} |S_j - E[S_j]| > \varepsilon\right] \leq \frac{1}{\varepsilon^2}Var[S_n],$$

donde, para $j \in \{1, \ldots, n\}$, $S_j = \sum_{i=1}^{j} X_i$.

Demostración

Supongamos primero que $E[X_k] = 0$ para cualquier $k \in \{1, \ldots, n\}$. Entonces también se tiene $E[S_k] = 0$ para cualquier $k \in \{1, \ldots, n\}$.

Sea $A = \left\{\omega \in \Omega : \max_{1\leq k\leq n} |S_k(\omega)| > \varepsilon\right\}$ y, para $k \in \{1,\ldots,n\}$:

$A_k = \left\{\omega \in A : \max_{1\leq j\leq k-1} |S_j(\omega)| \leq \varepsilon, |S_k(\omega)| > \varepsilon\right\}$,

donde $\max_{1\leq j\leq 0} |S_j(\omega)| \equiv 0$.

Entonces, los eventos $A_1,\ldots,A_n$ son mutuamente excluyentes y $A = \bigcup_{k=1}^n A_k$. Así que:

$E[S_n^2 I_A] = E[S_n^2 \sum_{k=1}^n I_{A_k}] = \sum_{k=1}^n E[S_n^2 I_{A_k}] = \sum_{k=1}^n E\left[(S_k + S_n - S_k)^2 I_{A_k}\right]$

$= \sum_{k=1}^n E\left[\left(S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\right) I_{A_k}\right]$

$= \sum_{k=1}^n E[S_k^2 I_{A_k}] + 2\sum_{k=1}^n E[S_k(S_n - S_k) I_{A_k}] + \sum_{k=1}^n E\left[(S_n - S_k)^2 I_{A_k}\right]$.

Pero, por la proposición 13.5 y el corolario 7.7, $S_k I_{A_k}$ y $S_n - S_k$ son independientes y tienen esperanza finita, de manera que, por la proposición 14.4, se tiene:

$E[S_k I_{A_k}(S_n - S_k)] = E[S_k I_{A_k}] E[S_n - S_k]$

$= E[S_k I_{A_k}] E[S_n - S_k] = 0$.

Por lo tanto:

$Var[S_n] = E[S_n^2] \geq E[S_n^2 I_A] = \sum_{k=1}^n E[S_k^2 I_{A_k}] + \sum_{k=1}^n E\left[(S_n - S_k)^2 I_{A_k}\right]$

$\geq \sum_{k=1}^n E[S_k^2 I_{A_k}] \geq \sum_{k=1}^n \varepsilon^2 E[I_{A_k}] = \varepsilon^2 \sum_{k=1}^n P(A_k) = \varepsilon^2 P(A)$

$= \varepsilon^2 P\left[\max_{1\leq j\leq n} |S_j - E[S_j]| > \varepsilon\right]$,

de lo cual se sigue el resultado.

Para el caso general, sea $Y_k = X_k - E[X_k]$ para $k \in \{1,\ldots,n\}$. Entonces, las variables aleatorias $Y_1,\ldots,Y_n$ son independientes, tienen varianza finita, $\sum_{i=1}^j Y_i = \sum_{i=1}^j (X_i - E[X_i])$ y $E[Y_j] = 0$ para cualquier $j \in \{1,\ldots,n\}$. De manera que si ε es cualquier número real positivo y $S_j = \sum_{i=1}^j X_i$ para cualquier $j \in \{1,\ldots,n\}$, entonces:

$P\left[\max_{1\leq j\leq n} |S_j - E[S_j]| > \varepsilon\right] = P\left[\max_{1\leq j\leq n} \left|\sum_{i=1}^j Y_i\right| > \varepsilon\right]$

$\leq \frac{1}{\varepsilon^2} Var\left[\sum_{i=1}^j Y_i\right] = \frac{1}{\varepsilon^2} Var[S_n]$.

■

TEOREMA 14.10 (**Ley fuerte de los grandes números (1) de Kolmogorov**). *Sea $X_1, X_2, \dots$ una sucesión de variables aleatorias, independientes, de varianza finita, esperanza nula y tales que $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, donde σ_n^2 es la varianza de X_n. Entonces:*

$\frac{X_1+\cdots+X_n}{n} \xrightarrow{c.s.} \mu$.

Demostración

Para cada $n \in \mathbb{N}$ sea $S_n = \sum_{k=1}^{n} X_k$ y, para cada $\varepsilon > 0$, sea:

$A_\varepsilon = \left\{\omega \in \Omega : \left|\frac{S_n(\omega)}{n}\right| > \varepsilon \text{ para una infinidad de valores de } n\right\}$.

Por la proposición 10.1, para probar el resultado basta con demostrar que $P(A_\varepsilon) = 0$ para cualquier $\varepsilon > 0$. Para esto definamos:

$B_{n,\varepsilon} = \left\{\omega \in \Omega : \left|\frac{S_k(\omega)}{k}\right| > \varepsilon \text{ para alguna } k \in \mathbb{N} \text{ tal que } 2^{n-1} < k \leq 2^n\right\}$.

Evidentemente se tiene:

$A_\varepsilon = \{\omega \in \Omega : \omega \in B_{n,\varepsilon} \text{ para una infinidad de valores de } n\}$.

De manera que, por el lema de Borel-Cantelli, para probar que $P(A_\varepsilon) = 0$ para cualquier $\varepsilon > 0$, basta con demostrar que $\sum_{n=1}^{\infty} P(B_{n,\varepsilon}) < \infty$ para cualquier $\varepsilon > 0$. Pero, utilizando la desigualdad de Kolmogorov, se tiene:

$P(B_{n,\varepsilon}) = P\left[\max_{2^{n-1}<k\leq 2^n} \left|\frac{S_k}{k}\right| > \varepsilon\right] = P\left[\max_{2^{n-1}<k\leq 2^n} |S_k| > k\varepsilon\right]$

$\leq P\left[\max_{2^{n-1}<k\leq 2^n} |S_k| > \varepsilon 2^{n-1}\right] \leq P\left[\max_{1\leq k\leq 2^n} |S_k| > \varepsilon 2^{n-1}\right]$

$\leq \frac{1}{\varepsilon^2 2^{2n-2}} Var\left[S_{2^n}\right] = \frac{4}{\varepsilon^2 2^{2n}} \sum_{k=1}^{2^n} \sigma_k^2$.

Así que:

$\sum_{n=1}^{\infty} P(B_{n,\varepsilon}) \leq \frac{4}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{2^{2n}} \sum_{k=1}^{2^n} \sigma_k^2 = \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{\{n\in\mathbb{N}:k\leq 2^n\}} \frac{1}{2^{2n}}$.

Sea ahora n_0 el más pequeño número natural tal que $k \leq 2^{n_0}$, entonces:

$\sum_{\{n\in\mathbb{N}:k\leq 2^n\}} \frac{1}{2^{2n}} = \sum_{n=n_0}^{\infty} \frac{1}{2^{2n}} = \frac{4}{2^{2n_0}} \leq \frac{4}{k^2}$.

Así que:

$\sum_{k=1}^{\infty} \sigma_k^2 \sum_{\{n\in\mathbb{N}:k\leq 2^n\}} \frac{1}{2^{2n}} \leq 4 \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$.

Por lo tanto:

$\sum_{n=1}^{\infty} P(B_{n,\varepsilon}) \leq \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{\{n \in \mathbb{N}: k \leq 2^n\}} \frac{1}{2^{2n}} < \infty.$

■

Corolario 14.10. *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes, de varianza finita y tales que $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, donde σ_n^2 es la varianza de X_n. Entonces:*

$P\left[\lim_{n \rightsquigarrow \infty} \frac{1}{n} \sum_{k=1}^{n} (X_k - E[X_k]) = 0\right] = 1.$

Para el caso en que las variables aleatorias $X_1, X_2, \ldots$ sean idénticamente distribuidas se cumple la ley fuerte con la única condición de que la esperanza común de $X_1, X_2, \ldots$ sea finita. La demostración de este resultado se debe también a Kolmogorov y el método de demostración es el de truncación, el cual fue utilizado en la demostración de la ley débil. Se requieren además algunos resultados previos, los cuales se exponen a continuación:

Lema 14.3. *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas de esperanza finita μ. Para $n \in \mathbb{N}$, definamos:*

$$Y_n = \begin{cases} X_n & \textit{si } |X_n| \leq n \\ 0 & \textit{en otro caso} \end{cases}$$

Entonces:

(i) $\lim_{n \rightsquigarrow \infty} E[Y_n] = \mu$.
(ii) *Y_n tiene varianza finita para cualquier $n \in \mathbb{N}$.*
(iii) $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, *en donde σ_n^2 es la varianza de Y_n.*
(iv) $P[\{\omega \in \Omega : \textit{existe } N(\omega) \textit{ tal que } Y_n(\omega) = X_n(\omega) \textit{ para cualquier } n \geq N(\omega)\}] = 1.$

Demostración

1. Se tiene:

$$F_{Yn}(x) = \begin{cases} 0 & \text{si } x < -n \\ P[-n \leq X_n \leq x] & \text{si } -n \leq x < 0 \\ P[|X_n| > n] + P[-n \leq X_n \leq x] & \text{si } 0 \leq x \leq n \\ 1 & \text{si } x > n \end{cases}$$

$$= \begin{cases} 0 & \text{si } x < -n \\ P[-n \leq X_n \leq x] & \text{si } -n \leq x < 0 \\ 1 - P[x < X_n \leq n] & \text{si } 0 \leq x \leq n \\ 1 & \text{si } x > n \end{cases}$$

Así que:

$E[Y_n] = \int_0^{\infty} [1 - F_{Yn}(x)]\, dx - \int_0^n F_{Yn}(-x) dx$

$= \int_0^n P[x < X_n \leq n]\, dx - \int_0^n P[-n \leq X_n \leq -x]\, dx$

$= \int_0^n P[x < X_1 \leq n]\, dx - \int_0^n P[-n \leq X_1 \leq -x]\, dx$

$= \int_0^n [1 - F_{X_1}(x)]\, dx - \int_0^n P[X_1 > n]\, dx - \int_0^n F_{X_1}(-x)dx. + \int_0^n P[X_1 < -n]\, dx$

$= \int_0^n [1 - F_{X_1}(x)]\, dx - \int_0^n F_{X_1}(-x)dx - nP[X_1 > n] + nP[X_1 < -n]$.

Por lo tanto, utilizando la proposición 14.9, $\lim_{n \rightsquigarrow \infty} E[Y_n] = E[X_1] = \mu$.

2. Para cualquier $n \in \mathbb{N}$, se tiene $|Y_n| \leq n$, así que Y_n tiene varianza finita.

3. $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} E[Y_n^2] = \sum_{n=1}^{\infty} \frac{1}{n^2} E\left[X_n^2 I_{[|X_n| \leq n]}\right]$

$= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{j=1}^{n} E\left[X_n^2 I_{[j-1<|X_n| \leq j]}\right]$

$= E\left[X_1^2 I_{[j-1<|X_1| \leq j]}\right] + \frac{1}{2^2}\left(E\left[X_1^2 I_{[j-1<|X_1| \leq j]}\right] + E\left[X_2^2 I_{[j-1<|X_2| \leq j]}\right]\right)$

$+ \ldots$

$= E\left[X_1^2 I_{[j-1<|X_1| \leq j]}\right]\left(1 + \frac{1}{2^2} + \cdots\right) + E\left[X_2^2 I_{[j-1<|X_2| \leq j]}\right]\left(\frac{1}{2^2} + \frac{1}{3^2} + \cdots\right)$

$+ \cdots$

$= \sum_{j=1}^{\infty} E\left[X_j^2 I_{[j-1<|X_j| \leq j]}\right] \sum_{n=j}^{\infty} \frac{1}{n^2}$.

Pero, para cualquier $j \in \{2, 3, \ldots\}$, se tiene:

$\sum_{n=j}^{\infty} \frac{1}{n^2} \leq \int_{j-1}^{\infty} \frac{1}{x^2} = \frac{1}{j-1} \leq \frac{2}{j}$.

Además, $\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \sum_{n=2}^{\infty} \frac{1}{n^2} \leq 2$.

Así que, $\sum_{n=j}^{\infty} \frac{1}{n^2} \leq \frac{2}{j}$ para cualquier $j \in \mathbb{N}$.

Además, tomando en cuenta que $X_1, X_2, \ldots$ tienen la misma distribución:

$E\left[X_j^2 I_{[j-1<|X_j| \leq j]}\right] \leq jE\left[|X_j| I_{[j-1<|X_j| \leq j]}\right] = jE\left[|X_1| I_{[j-1<|X_1| \leq j]}\right]$.

Por lo tanto:

$\sum_{j=1}^{\infty} E\left[X_j^2 I_{[j-1<|X_j| \leq j]}\right] \sum_{n=j}^{\infty} \frac{1}{n^2} \leq \sum_{j=1}^{\infty} jE\left[|X_1| I_{[j-1<|X_1| \leq j]}\right] \frac{4}{j}$

$= 4 \sum_{j=1}^{\infty} E\left[|X_1| I_{[j-1<|X_1| \leq j]}\right]$.

Sea ahora $Z_n = \sum_{j=1}^{n} |X_1| I_{[j-1<|X_1| \leq j]}$, entonces la sucesión de variables aleatorias $Z_1, Z_2, \ldots$ es no decreciente y $\lim_{n \rightsquigarrow \infty} Z_n(\omega) = |X_1(\omega)|$ para cualquier $\omega \in \Omega$, así que por el teorema de la convergencia monótona:

$\sum_{j=1}^{\infty} E\left[|X_1| I_{[j-1<|X_1| \leq j]}\right] = \lim_{n \rightsquigarrow \infty} E[Z_n] = E[|X_1|] < \infty$,

de lo cual se sigue que $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$.

4. $P[Y_n \neq X_n] = P[|X_n| > n] = P[|X_1| > n]$.

De manera que, utilizando la proposición 8.1:

$\sum_{n=1}^{\infty} P[Y_n \neq X_n] = \sum_{n=1}^{\infty} P[|X_1| > n] \leq \sum_{n=1}^{\infty} P[|X_1| \geq n] < \infty.$

Así que, por el lema de Borel-Cantelli, si:

$A = \{\omega \in \Omega : Y_n(\omega) \neq X_n(\omega)$ para una infinidad de valores de $n\}$,

entonces $P(A) = 0$.

Sea ahora:

$B = \{\omega \in \Omega :$ existe $N(\omega)$ talque $Y_n(\omega) = X_n(\omega)$ para cualquier $n \geq N(\omega)\}$.

Entonces, $B \supset A^c$, así que, $P(B) \geq P(A) = 1$. ∎

Corolario 14.11. *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas de esperanza finita. Para $n \in \mathbb{N}$, definamos:*

$$Y_n = \begin{cases} X_n & si\ |X_n| \leq n \\ 0 & en\ otro\ caso \end{cases}.$$

Entonces:

$P\left[\left\{\omega \in \Omega : \text{lím}_{n\rightsquigarrow\infty} \frac{1}{n}\sum_{k=1}^{n}[X_k(\omega) - Y_k(\omega)] = 0\right\}\right] = 1.$

Demostración

Por la parte *iv* del lema 14.3, si:

$B = \{\omega \in \Omega :$ existe $N(\omega)$ talque $Y_n(\omega) = X_n(\omega)$ para cualquier $n \geq N(\omega)\}$,

entonces $P(B) = 1$.

Pero si $\omega \in B$, entonces existe $N(\omega)$ tal que $X_n(\omega) - Y_n(\omega) = 0$ para cualquier $n \geq N(\omega)$, así que:

$\text{lím}_{n\rightsquigarrow\infty} \frac{1}{n}\sum_{k=1}^{n}[X_k(\omega) - Y_k(\omega)] = 0.$

∎

Lema 14.4. *Sea (x_n) una sucesión convergente de números reales y sea $x = \text{lím}_{n\rightsquigarrow\infty} x_n$. Entonces la sucesión $z_n = \frac{1}{n}\sum_{k=1}^{n} x_k$ es convergente y $\text{lím}_{n\rightsquigarrow\infty} z_n = x$.*

Demostración

Sea $M > 0$ tal que $|x - x_n| \leq M$ para cualquier $n \in \mathbb{N}$.

Dada $\varepsilon > 0$, sea $m \in \mathbb{N}$ tal que $|x - x_n| < \frac{\varepsilon}{2}$ para cualquier $n \geq m$.

Entonces, para $n > \text{máx}\left\{m, \frac{2mM}{\varepsilon}\right\}$, se tiene:

$|z_n - x| = \left|\frac{1}{n}\sum_{k=1}^{n} x_k - x\right| = \left|\frac{1}{n}\sum_{k=1}^{n}(x_k - x)\right| \leq \frac{1}{n}\sum_{k=1}^{n}|x_k - x|$

$= \frac{1}{n}\sum_{k=1}^{m}|x_k - x| + \frac{1}{n}\sum_{k=m+1}^{n}|x_k - x|$

$\leq \frac{mM}{n} + \frac{(n-m)\varepsilon}{2n} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$

lo cual significa que $\text{lím}_{n\rightsquigarrow\infty} z_n = x$.

■

Teorema 14.11 (**Ley fuerte de los grandes números (2) de Kolmogorov**). *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ. Entonces:*

$\frac{X_1+\cdots+X_n}{n} \xrightarrow{c.s.} \mu.$

Demostración

Para cada $n \in \mathbb{N}$, sea:

$Y_n = \begin{cases} X_n & \text{si } |X_n| \leq n \\ 0 & \text{en otro caso} \end{cases}$

Por el lema 14.3, las variables aleatorias $Y_1, Y_2, \ldots$ tienen esperanza finita, $\text{lím}_{n\rightsquigarrow\infty} E[Y_n] = \mu$ y $\sum_{n=1}^{\infty}\frac{\sigma_n^2}{n^2} < \infty$, donde σ_n^2 es la varianza de Y_n. De manera que, por el lema 14.4 y el corolario 14.10, se tiene:

$\text{lím}\,\frac{1}{n}\sum_{k=1}^{n} E[Y_k] = \mu$

$P\left[\left\{\omega \in \Omega : \text{lím}_{n\rightsquigarrow\infty}\frac{1}{n}\sum_{k=1}^{n}(Y_k(\omega) - E[Y_k]) = 0\right\}\right] = 1,$

de lo cual se obtiene:

$P\left[\left\{\omega \in \Omega : \text{lím}_{n\rightsquigarrow\infty}\frac{1}{n}\sum_{k=1}^{n} Y_k(\omega) = \mu\right\}\right] = 1.$

Además, por el corolario 14.11:

$P\left[\left\{\omega \in \Omega : \text{lím}_{n\rightsquigarrow\infty}\frac{1}{n}\sum_{k=1}^{n}[X_k(\omega) - Y_k(\omega)] = 0\right\}\right] = 1,$

de lo cual se obtiene el resultado.

■

CAPÍTULO 15

CONSTRUCCIÓN DE ESPACIOS DE PROBABILIDAD

15.1. Introducción

La Teoría de la Probabilidad surgió del planteamiento de problemas teóricos, los cuales provenían de algún problema práctico. El problema de la división de apuestas, por ejemplo, surgió al buscar determinar cómo deberían de repartirse las apuestas en un juego de azar que se interrumpe antes de que alguno de los participantes gane el juego de acuerdo con las reglas establecidas. Sin embargo, como lo mencionamos en el capítulo anterior, una vez que la Teoría de la Probabilidad se formula en forma axiomática, los elementos que la componen no requieren de una interpretación práctica. Más aún, en la formulación axiomática, no es admisible definir un concepto en términos de un determinado fenómeno aleatorio, o demostrar algún teorema utilizando propiedades de un fenómeno aleatorio. Es posible que se utilice la Teoría de la Probabilidad para modelar algún fenómeno aleatorio y que los conceptos que se introduzcan o las propiedades que se demuestren provengan de las características de dicho fenómeno, o que estemos interesados en estudiar las propiedades del fenómeno en consideración basándonos en el modelo matemático, pero, si bien éste puede ser el caso, los elementos mismos del fenómeno o sus propiedades no forman parte del cuerpo teórico; lo que se observa del fenómeno podría motivar introducir algún concepto o buscar algún resultado dentro del modelo matemático que nos ayude a entender el fenómeno o darnos una idea de sus propiedades, pero las observaciones mismas no forman parte del modelo. Por ejemplo, el lanzamiento de un dado n veces consecutivas lo podemos modelar utilizando como espacio muestral al conjunto de todos los posibles resultados de los n lanzamientos, es decir, al conjunto Ω formado por todas las colecciones ordenadas de n números naturales que pertenecen al conjunto $\{1,2,3,4,5,6\}$; la familia de eventos la podemos considerar como el conjunto potencia de Ω, es decir la familia formada por todos los subconjuntos de Ω; finalmente, como medida de probabilidad podemos tomar la que asigna a cada subconjunto de Ω el cociente que resulta de dividir el número de elementos de ese subconjunto entre 6^n, el cual es el número de elementos de Ω. Tendríamos así definido nuestro espacio de probabilidad, el cual podemos pensarlo como modelo matemático del lanzamiento de n dados en forma consecutiva, pero el modelo mismo es una abstracción, no requerimos de referirnos al lanzamiento del dado para definirlo. Podríamos decir: un ejemplo de espacio de probabilidad es el siguiente: Definamos $\Omega = \{1,2,3,4,5,6\}$, $\Im$ como la familia de todos los subconjuntos de Ω y como medida de probabilidad tomemos a la función $P : \Im \to \mathbb{R}$ definida, para cada $B \in \Im$, mediante la relación $P(B) = \sum_{\{\omega \in B\}} p(\omega)$, donde $p(\omega) = \frac{1}{6^n}$ para cualquier $\omega \in \Omega$. El espacio de probabilidad $(\Omega, \Im, P)$ así definido está formado por elementos matemáticos abstractos que se tendrían que tratar como independientes de cualquier fenómeno aleatorio.

Como ya lo mencionamos, en algunos problemas de probabilidad se hace referencia al concepto de experimento aleatorio y se construye un espacio de probabilidad ad hoc para ese experimento, de manera similar a como lo hicimos con el lanzamiento del dado, pero nuevamente, el espacio de probabilidad que se construye tiene que tratarse como independiente de cualquier fenómeno aleatorio.

Un caso particular de experimento aleatorio es lo que se conoce como ensayo de Bernoulli, el cual se define como un experimento aleatorio que admite únicamente dos posibles resultados, a uno de los cuales se le llama éxito y al otro fracaso. Utilizando este concepto, podemos, por ejemplo, definir algunas variables aleatorias de interés o plantearnos algunos problemas de probabilidad que historicamente fueron importantes para el desarrollo de la Teoría de la Probabilidad.

Por, ejemplo, podemos considerar al experimento aleatorio que consiste en la realización consecutiva de n ensayos de Bernoulli, cada uno de ellos independiente de los demás y tal que la probabilidad de obtener éxito es igual a un número $p \in [0, 1]$, el cual es el mismo para cada uno de los ensayos. Si definimos X como el número de éxitos que se obtienen al realizar los n ensayos, X resulta ser una variable aleatoria con distribución binomial de parámetros n y p, lo cual podemos mostrar sin necesidad de definir formalmente algún espacio de probabilidad.

Como otro ejemplo, podemos considerar al experimento aleatorio que consiste en la realización consecutiva de una infinidad de ensayos de Bernoulli, cada uno de ellos independiente de los demás y tal que la probabilidad de obtener éxito es igual a un número $p \in [0, 1]$, el cual es el mismo para cada uno de los ensayos. Si definimos Y como el número de fracasos que se obtienen antes de obtener éxito por primera vez al realizar el experimento, X resulta ser una variable aleatoria con distribución geométrica de parámetro p, lo cual, como en el caso anterior, podemos mostrar sin necesidad de definir formalmente algún espacio de probabilidad.

Como tercer ejemplo, podemos, nuevamente, considerar al experimento aleatorio que consiste en la realización consecutiva de una infinidad de ensayos de Bernoulli, cada uno de ellos independiente de los demás y tal que la probabilidad de obtener éxito es igual a un número $p \in [0, 1]$, el cual es el mismo para cada uno de los ensayos. Pero esta vez podemos preguntarnos por la probabilidad de obtener un número finito de éxitos al realizar el experimento.

En el caso del primer ejemplo, no hay problema para tratarlo sin necesidad de recurrir a la formulación axiomática que tratamos en el capítulo anterior. Pero, en el segundo ejemplo, y sobre todo en el tercero, podemos observar un problema si no se recurre a una formulación matemática (abstracta). El problema consiste en que la realización de una infinidad de ensayos de Bernoulli es imposible. En el segundo ejemplo podría argumentarse que no hay tal problema ya que casi con seguridad se obtendría el primer éxito en un número finito de ensayos; sin embargo, sí hay problema, ya que siendo independiente cada ensayo de los demás, en cualquiera de ellos es posible que el resultado sea fracaso; de manera que es posible que no podamos determinar el número de fracasos que se obtienen antes de obtener éxito por primera vez. Podríamos obtener la distribución de X imaginando que es posible la realización del experimento, pero estaríamos partiendo de algo que es falso. En el tercer ejemplo es completamente claro que no es posible determinar si se obtiene un número finito de éxitos al realizar el experimento ya que para determinarlo es necesario conocer la infinidad de resultados que se obtienen, lo cual no es posible.

Lo anterior muestra uno de los problemas a los que se enfrenta uno al no contar con un modelo matemático abstracto que permita formular los problemas de una manera distinta, dentro de un marco teórico donde no haya necesidad de realizar experimentos. Esto tiene relación con el planteamiento de Poincaré cuando, en el año 1896, decía: "No se puede dar una definición satisfactoria de la probabilidad." y agregaba: "La definición completa de la probabilidad es una especie de petición de principio... deberemos, en cada aplicación, hacer convenciones." Obsérvese que Poincaré incluye la frase "en cada aplicación", lo cual nos dice que un problema de probabilidad lo pensaba vinculado a un determinado problema práctico. Es decir, la teoría y la aplicación estaban mezcladas en una sola cosa. Esto ilustra una de las razones por las cuales el Cálculo de Probabilidades no era considerado en esa época como una rama de las Matemáticas, sino de la Física, tal como lo expresó Hilbert en el año 1900: "Las investigaciones sobre los principios fundamentales de la geometría nos conducen a plantear este problema: Tratar con base en ese modelo las ramas de la Física donde las Matemáticas juegan actualmente un papel preponderante; esas ramas de la ciencia son, antes que cualesquiera otras, el Cálculo de Probabilidades y la Mecánica."

Con la formulación axiomática de la Teoría de la Probabilidad, la teoría y las aplicaciones quedan separadas, aunque entrelazadas de alguna manera. La teoría puede ser desarrollada independientemente de las aplicaciones que se hagan de ella, sin quedar éstas por fuera completamente ya que son las aplicaciones las que hacen surgir y alimentan las teorías, complementándose unas con la otras.

Pasemos a ver de qué manera se resuelve el problema que planteamos en los ejemplos anteriores al formularlos dentro del marco teórico que tenemos desarrollado.

En primer lugar, en lugar de hablar de un ensayo de Bernoulli, lo que se hace es introducir lo que se conoce como distribución Bernoulli. Para esto se representa un éxito con el número 1 y un fracaso con el número 0. Entonces decimos que una variable aleatoria X tiene distribución Bernoulli, con parámetro $p \in [0,1]$, si $P[X=1]=p$ y $P[X=1]=1-p$.

En lugar de considerar al experimento aleatorio que consiste en la realización consecutiva de n ensayos de Bernoulli, cada uno de ellos independiente de los demás y tal que la probabilidad de obtener éxito es igual a un número $p \in [0,1]$, lo que se hace es construir un espacio de probabilidad $(\Omega, \Im, P)$ en el cual se puedan definir n variables aleatorias independientes, cada una con distribución Bernoulli de parámetro p. Este espacio ya lo definimos en el capítulo anterior:

Sea $\Omega_n = \{(s_1, s_2, \ldots, s_n) : s_j \in \{0,1\}$ para cualquier $j \in \{1, 2, \ldots, n\}\}$ y, para cada $\omega = (s_1, \ldots, s_n) \in \Omega_n$ y cada subconjunto A de Ω, definamos:

$p_n(\omega) = \prod_{j=1}^{n} [ps_j + (1-p)(1-s_j)]$,

$P(A) = \sum_{\{\omega \in A\}} p_n(\omega)$.

Como $\sum_{\{\omega \in \Omega_n\}} p_n(\omega) = 1$, P es una medida de probabilidad definida sobre el conjunto potencia de Ω.

Definamos ahora, para cada $j \in \{1, 2, \ldots, n\}$, $X_j : \Omega_n \to \mathbb{R}$ mediante la relación:

$X_j((s_1, s_2, \ldots, s_n)) = s_j$.

Entonces, para cualquier $k \in \{1, 2, \ldots, n\}$, dados $r_1, r_2, \ldots, r_k \in \{0,1\}$, se tiene:

$P\left(\bigcap_{j=1}^{k} [X_j = r_j]\right) = \sum_{\{(s_1,s_2,\ldots,s_n)\in\Omega_n : s_j = r_j \text{ para cualquier } j\in\{1,\ldots,k\}\}} \prod_{j=1}^{n} [ps_j + (1-p)(1-s_j)]$

$= \left(\prod_{j=1}^{k} [pr_j + (1-p)(1-r_j)]\right)\cdot$

$\left(\sum_{\{(s_{k+1},\ldots,s_n) : s_j\in\{0,1\} \text{ para cualquier } j\in\{k+1,\ldots,n\}\}} \prod_{j=k+1}^{n} [ps_j + (1-p)(1-s_j)]\right)$

$= \prod_{j=1}^{k} [pr_j + (1-p)(1-r_j)]$.

Así que, si $r \in \{0,1\}$, se tiene:

$P[X_k = r]$

$= \sum_{\{(s_1,s_2,\ldots,s_{k-1}) : s_j\in\{0,1\} \text{ para cualquier } j\in\{1,\ldots,k-1\}\}} P[X_1 = s_1, X_2 = s_2, \ldots, X_{k-1} = s_{j-1}, X_k = r]$

$= \sum_{\{(s_1,s_2,\ldots,s_{k-1}) : s_j\in\{0,1\} \text{ para cualquier } j\in\{1,\ldots,k-1\}\}} [pr + (1-p)(1-r)] \prod_{j=1}^{k-1} [pr_j + (1-p)(1-r_j)]$

$= [pr + (1-p)(1-r)] \sum_{\{(s_1,s_2,\ldots,s_{k-1}) : s_j\in\{0,1\} \text{ para cualquier } j\in\{1,\ldots,k-1\}\}} \prod_{j=1}^{k-1} [pr_j + (1-p)(1-r_j)]$

$= pr + (1-p)(1-r)$.

Por lo tanto, para cualquier $k \in \{1, 2, \ldots, n\}$, X_k tiene distribución Bernoulli con parámetro p.

Además, para cualquier $(s_1, s_2, \ldots, s_n) \in \Omega_n$, se tiene:

$P\left(\bigcap_{j=1}^{n} [X_j = s_j]\right) = \prod_{j=1}^{n} [ps_j + (1-p)(1-s_j)] = \prod_{j=1}^{n} P[X_j = s_j]$.

Así que las variables aleatorias $X_1, X_2, \ldots, X_n$ son independientes.

Ahora, si definimos $X : \Omega_n \to \mathbb{R}$ mediante la relación $X = \sum_{j=1}^{n} X_j$, X tiene distribución binomial con parámetros n y p.

15.2. Funciones de distribución como medidas

En este capítulo veremos cómo, partiendo de la función de distribución conjunta de un vector aleatorio $(X_1, \ldots, X_n)$ con valores en $\mathbb{R}^n$ podemos obtener la medida de probabilidad $\mu_{X_1,\ldots,X_n}$ asociada con el vector aleatorio.

DEFINICIÓN 15.1. *Diremos que una función $F : \mathbb{R} \mapsto \mathbb{R}$ es una función de distribución finita en 1 variable si satisface las siguientes propiedades:*

(i) *F es una función no decreciente y continua por la derecha.*
(ii) $\lim_{x \rightsquigarrow \infty} F_X(x) < \infty$
(iii) $\lim_{x \rightsquigarrow -\infty} F_X(x) = 0$

DEFINICIÓN 15.2. *Para $n \in \{2, 3, \ldots\}$, diremos que una función $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una función de distribución finita en n variables si satisface las siguientes propiedades:*

(i) $\sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1,\cdots,x_n) \in S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}\right\}} F(x_1, \cdots, x_n) \geq 0$
para cualquier rectángulo $(a_1, b_1] \times \cdots \times (a_n, b_n]$.
(ii) $\lim_{m \rightsquigarrow 0} F\left(x_1 + \delta_1^{(m)}, \cdots, x_n + \delta_n^{(m)}\right) = F(x_1, \cdots, x_n)$
para cualquier vector $(x_1, \cdots, x_n) \in \mathbb{R}^n$ y cualquier sucesión $\left(\left(\delta_1^{(m)}, \cdots, \delta_n^{(m)}\right)\right)_{m \in \mathbb{N}}$ que converja al vector $\overline{0} \in \mathbb{R}^n$ y tal que $\delta_1^{(m)}, \cdots, \delta_n^{(m)}$ sean números reales positivos.
(iii) $\lim_{x \rightsquigarrow -\infty} F(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_n) = 0$
para cualquier $(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \in \mathbb{R}^{n-1}$.
(iv) *Para cada $(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \in \mathbb{R}^{n-1}$, el límite*
$\lim_{x \rightsquigarrow \infty} F(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_n)$
existe y la función $G : \mathbb{R}^{n-1} \mapsto \mathbb{R}$ definida por:
$G(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) = \lim_{x \rightsquigarrow \infty} F(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_n)$
es una función de distribución finita en $n-1$ variables.

Cuando $\lim_{(x_1,\cdots,x_n) \rightsquigarrow (\infty,\ldots,\infty)} F(x_1, \cdots, x_n) = 1$, diremos simplemente que F es una función de distribución en n variables.

Si $a, b \in \overline{\mathbb{R}}$ y $a \leq b$, entonces definimos $(a, b|$ de la siguiente manera:

$$(a, b| = \begin{cases} (a, b] & \text{si } b \in \mathbb{R} \\ (a, b) & \text{si } b = \infty \end{cases}$$

Si $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una función de distribución finita y

$(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \in \mathbb{R}^{n-1}$, definimos:

$F(x_1, \ldots, x_{j-1}, \infty, x_{j+1}, \ldots, x_n) = \text{lím}_{x \rightsquigarrow \infty} F(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_n)$.

$F(x_1, \ldots, x_{j-1}, -\infty, x_{j+1}, \ldots, x_n) = 0$.

Con estas convenciones, se tiene que:

$\sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1,\cdots,x_n) \in S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}\right\}} F_{X_1,\ldots,X_n}(x_1, \cdots, x_n) \geq 0$

para cualquier rectángulo $(a_1, b_1| \times \cdots \times (a_n, b_n|$.

Definición 15.3. *Si $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una función de distribución finita y $R = (a_1, b_1| \times \cdots \times (a_n, b_n|$ es un rectángulo en $\mathbb{R}^n$, definimos $\mu_F(R)$ de la siguiente manera:*

$\mu_F(R) = \sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1,\cdots,x_n) \in S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}\right\}} F(x_1, \cdots, x_n)$.

Lema 15.1. *Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución finita y $R = (a_1, b_1| \times \cdots \times (a_n, b_n|$ un rectángulo en $\mathbb{R}^n$. Para cada intervalo $(a_i, b_i|$ consideremos una partición:*

$P_i = \left\{ a_i = c_0^{(i)} < c_1^{(i)} < \cdots < c_{m_i}^{(i)} = b_i \right\}$.

Entonces:

$\mu_F(R) = \sum_{j_i \in \{1,\ldots,m_i\}} \mu_F\left(R_{\left(c^{(1)}_{j_1-1}, c^{(1)}_{j_1}, \ldots, c^{(n)}_{j_n-1}, c^{(n)}_{j_n}\right)} \right)$.

Demostración

Las particiones P_i parten el rectángulo R en $m_1 \cdots m_n$ rectángulos de la forma $\left(c^{(1)}_{j_1-1}, c^{(1)}_{j_1}\right| \times \cdots \times \left(c^{(n)}_{j_n-1}, c^{(n)}_{j_n}\right|$ y se tiene: $(a_1, b_1| \times \cdots \times (a_n, b_n| = \bigcup_{j_i \in \{1,\ldots,m_i\}} \left(c^{(1)}_{j_1-1}, c^{(1)}_{j_1}\right| \times \cdots \times \left(c^{(n)}_{j_n-1}, c^{(n)}_{j_n}\right|$.

Denotemos por $V_{\left(c^{(1)}_{j_1-1}, c^{(1)}_{j_1}, \ldots, c^{(n)}_{j_n-1}, c^{(n)}_{j_n}\right)}$ al conjunto de vértices del rectángulo $\left(c^{(1)}_{j_1-1}, c^{(1)}_{j_1}\right| \times \cdots \times \left(c^{(n)}_{j_n-1}, c^{(n)}_{j_n}\right|$, por S a la sumatoria:

$$\sum_{j_i \in \{1,\ldots,m_i\}} \left[\sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1,\cdots,x_n) \in S^{(k)}_{\left(c^{(1)}_{j_1-1}, c^{(1)}_{j_1}, \ldots, c^{(n)}_{j_n-1}, c^{(n)}_{j_n}\right)}\right\}} F(x_1, \cdots, x_n) \right],$$

y por S' a la sumatoria:

$\sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1,\cdots,x_n) \in S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}\right\}} F(x_1, \cdots, x_n)$.

Sea $(x_1, \cdots, x_n) \in V_{\left(c^{(1)}_{j_1-1}, c^{(1)}_{j_1}, \ldots, c^{(n)}_{j_n-1}, c^{(n)}_{j_n}\right)}$, entonces cada coordenada x_i es un elemento de la partición P_i.

Si $x_i = c_j^{(i)} \in P_i - \{a_i, b_i\}$, entonces $(x_1, \cdots, x_n)$ es vértice de un rectángulo cuyo i-ésimo lado es $\left(c_{j-1}^{(i)}, c_j^{(i)}\right|$ y también es vértice de un rectángulo cuyo i-ésimo lado es $\left(c_j^{(i)}, c_{j+1}^{(i)}\right|$, de manera que $F(x_1, \cdots, x_n)$ aparecerá en la sumatoria S dos veces, una con signo positivo y otra con signo negativo, cancelándose.

Por lo tanto, los únicos términos de la sumatoria S, que no se anulan, son aquellos para los cuales $x_i \in \{a_i, b_i\}$ para toda $i \in \{1, \ldots, n\}$, es decir, $(x_1, \cdots, x_n) \in S_k$ para alguna $k \in \{0, \ldots, n\}$.

Cuando $x_i = a_i$, entonces el punto $(x_1, \cdots, x_n)$ es vértice de un rectángulo cuyo i-ésimo lado es $\left(a_i, c_1^{(i)}\right|$, mientras que cuando $x_i = b_i$, entonces el punto $(x_1, \cdots, x_n)$ es vértice de un rectángulo cuyo i-ésimo lado es $\left(c_{m_i-1}^{(i)}, b_i\right|$. Por lo tanto, si $(x_1, \cdots, x_n) \in S_k$, entonces $(x_1, \cdots, x_n) \in S^{(k)}_{\left(c_{j_1-1}^{(1)}, c_{j_1}^{(1)}, \ldots, c_{j_n-1}^{(n)}, c_{j_n}^{(n)}\right)}$ para alguna colección $(j_1, \ldots, j_n)$, así que $F(x_1, \cdots, x_n)$ aparece en la sumatoria S y en la sumatoria S' con el mismo signo. Es decir, $S = S'$.

■

Proposición 15.1. *Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución finita, $R = (a_1, b_1| \times \cdots \times (a_n, b_n|$ un rectángulo en $\mathbb{R}^n$ y $R^{(j)} = \left(a_1^{(j)}, b_1^{(j)}\right| \times \cdots \times \left(a_n^{(j)}, b_n^{(j)}\right|$ una colección finita de rectángulos en $\mathbb{R}^n$, ajenos por parejas, tal que $R = \bigcup_{j=1}^m R^{(j)}$, entonces:*

$\mu_F(R) = \sum_{j=1}^m \mu_F\left(R^{(j)}\right)$.

Demostración

Para cada $i \in \{1, \ldots, n\}$, los puntos $a_i^{(1)}, b_i^{(1)}, \ldots, a_i^{(m)}, b_i^{(m)}$ constituyen una partición del intervalo $(a_i, b_i|$.

Este conjunto de particiones parte cada rectángulo $R^{(j)}$ en subrectángulos $R_1^{(j)}, \ldots, R_{i_j}^{(j)}$. Por el lema anterior, se tiene:

$\mu_F\left(R^{(j)}\right) = \sum_{k=1}^{i_j} \mu_F\left(R_k^{(j)}\right)$.

Además, $R = \bigcup_{j=1}^m \bigcup_{k=1}^{i_j} R_k^{(j)}$.

Así que, nuevamente, por el lema:

$\mu_F(R) = \sum_{j=1}^m \sum_{k=1}^{i_j} \mu_F\left(R_k^{(j)}\right) = \sum_{j=1}^m \mu_F\left(R^{(j)}\right)$.

■

Corolario 15.1. *Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una función de distribución finita, $R = (a_1, b_1| \times \cdots \times (a_n, b_n|$ un rectángulo en $\mathbb{R}^n$ y $R^{(j)} = \left(a_1^{(j)}, b_1^{(j)}\right| \times \cdots \times \left(a_n^{(j)}, b_n^{(j)}\right|$ una colección finita de rectángulos en $\mathbb{R}^n$ tal que $R \subset \bigcup_{j=1}^m R^{(j)}$, entonces:*

$\mu_F(R) \leq \sum_{j=1}^{m} \mu_F\left(R^{(j)}\right)$.

Demostración

Para cada $i \in \{1, \ldots, n\}$, los puntos $a_i, b_i, a_i^{(1)}, b_i^{(1)}, \ldots, a_i^{(m)}, b_i^{(m)}$ constituyen una partición de un intervalo $(c_i, d_i|$.

Este conjunto de particiones parte cada rectángulo $R^{(j)}$ en subrectángulos $R_1^{(j)}, \ldots, R_{i_j}^{(j)}$. Por el lema anterior, se tiene:

$\mu_F\left(R^{(j)}\right) = \sum_{k=1}^{i_j} \mu_F\left(R_k^{(j)}\right)$.

El conjunto de particiones definido antes también parte el rectángulo R en subrectángulos $R_1, \ldots, R_i$, así que, nuevamente por el lema, se tiene:

$\mu_F(R) = \sum_{k=1}^{i} \mu_F(R_k)$.

Por otra parte, como $R \subset \bigcup_{j=1}^{m} R^{(j)}$, cada rectángulo R_k coincide con un rectángulo $R_{k'}^{(j)}$ para alguna j y alguna k', por lo tanto:

$\mu_F(R) = \sum_{k=1}^{i} \mu_F(R_k)$

$\leq \sum_{j=1}^{m} \sum_{k=1}^{i_j} \mu_F\left(R_k^{(j)}\right) = \sum_{j=1}^{m} \mu_F\left(R^{(j)}\right)$.

■

Proposición 15.2. *Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución finita y:*

$R_1, \ldots, R_k$ y $R^{(1)}, \ldots, R^{(m)}$ dos colecciones finitas de rectángulos en $\mathbb{R}^n$, todos de la forma:

$\left(a_1^{(1)}, b_1^{(1)}\right| \times \cdots \times \left(a_n^{(1)}, b_n^{(1)}\right|$

y tales que $R_1, \ldots, R_k$ son ajenos por parejas, $R^{(1)}, \ldots, R^{(m)}$ son ajenos por parejas y $\bigcup_{i=1}^{k} R_i = \bigcup_{j=1}^{m} R^{(j)}$.

Entonces:

$\sum_{i=1}^{k} \mu_F(R_i) = \sum_{j=1}^{m} \mu_F\left(R^{(j)}\right)$.

Demostración

Para cada $i \in \{1, \ldots, k\}$ y $j \in \{1, \ldots, m\}$, definamos $R_i^{(j)} = R_i \cap R^{(j)}$. Entonces, como $\bigcup_{i=1}^{k} R_i = \bigcup_{j=1}^{m} R^{(j)}$, se tiene $R_i = \bigcup_{j=1}^{m} R_i^{(j)}$ y $R^{(j)} = \bigcup_{i=1}^{k} R_i^{(j)}$, así que:

$\mu_F(R_i) = \sum_{j=1}^{m} \mu_F\left(R_i^{(j)}\right)$,

$\mu_F\left(R^{(j)}\right) = \sum_{i=1}^{k} \mu_F\left(R_i^{(j)}\right)$.

Por lo tanto:

$\sum_{i=1}^{k} \mu_F(R_i) = \sum_{i=1}^{k} \sum_{j=1}^{m} \mu_F\left(R_i^{(j)}\right)$

$= \sum_{j=1}^{m} \sum_{i=1}^{k} \mu_F\left(R_i^{(j)}\right) = \sum_{j=1}^{m} \mu_F\left(R^{(j)}\right).$

$\sum_{j=1}^{m} \mu_F\left(R^{(j)}\right) \leq \sum_{j=1}^{m} \sum_{k=1}^{\infty} \mu_F\left(R_k^{(j)}\right)$

$= \sum_{k=1}^{\infty} \sum_{j=1}^{m} \mu_F\left(R_k^{(j)}\right) = \sum_{k=1}^{\infty} \mu_F(R_k) = \sum_{i=1}^{\infty} \mu_F(A_i).$

■

TEOREMA 15.1. *Sea* $F : \mathbb{R}^n \mapsto \mathbb{R}$ *una función de distribución finita,* $R = (a_1, b_1| \times \cdots \times (a_n, b_n|$ *un rectángulo en* $\mathbb{R}^n$ *y* $R^{(i)} = \left(a_1^{(i)}, b_1^{(i)}\right| \times \cdots \times \left(a_n^{(i)}, b_n^{(i)}\right|$ *una colección infinita de rectángulos en* $\mathbb{R}^n$ *tal que* $R \subset \bigcup_{i=1}^{\infty} R^{(i)}$*, entonces:*

$\mu_F(R) \leq \sum_{i=1}^{\infty} \mu_F\left(R^{(i)}\right).$

Demostración

Para cada $i \in \mathbb{N}$ y cada $\delta_i > 0$, definamos, para $k \in \{1, \ldots, n\}$:

$$d_k^{\delta_i} = \begin{cases} b_k^{(i)} + \delta_i & \text{si } b_k^{(i)} \in \mathbb{R} \\ b_k^{(i)} & \text{si } b_k^{(i)} = \infty \end{cases}$$

Consideremos el rectángulo:

$R_{\delta_i} = \left(a_1^{(i)}, d_1^{\delta_i}\right| \times \cdots \times \left(a_n^{(i)}, d_n^{\delta_i}\right|,$

el cual contiene a $R^{(i)}$. Se tiene entonces:

$\lim_{\delta_i \rightsquigarrow 0} \mu_F(R_{\delta_i})$

$= \lim_{\delta_i \rightsquigarrow 0} \sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1,\cdots,x_n) \in S^{(k)}_{\left(a_1, d_1^{\delta_i}, \ldots, a_n, d_n^{\delta_i}\right)}\right\}} F(x_1, \cdots, x_n)$

$= \sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1,\cdots,x_n) \in S^{(k)}_{(a_1, b_1, \ldots, a_n, b_n)}\right\}} F(x_1, \cdots, x_n) = \mu_F\left(R^{(i)}\right).$

Dada $\varepsilon > 0$, existe entonces $\delta_i > 0$ tal que:

$\mu_F(R_{\delta_i}) - \mu_F\left(R^{(i)}\right) < \frac{\varepsilon}{2^i}.$

Por otra parte, si $\delta > 0$, definamos, para $k \in \{1, \ldots, n\}$:

$$c_{k,\delta} = \begin{cases} a_k + \delta & \text{si } a_k \in \mathbb{R} \\ -\frac{1}{\delta} & \text{si } a_k = -\infty \end{cases}$$

$d_{k,\delta} = \begin{cases} b_k & \text{si } b_k \in \mathbb{R} \\ \frac{1}{\delta} & \text{si } b_k = \infty \end{cases}$

Consideremos el rectángulo:

$R_\delta = (c_{1,\delta}, d_{1,\delta}] \times \cdots \times (c_{n,\delta}, d_{n,\delta}]$.

Entonces:

$\lim_{\delta \rightsquigarrow 0} \mu_F(R_\delta)$

$= \lim_{\delta \rightsquigarrow 0} \sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1,\cdots,x_n) \in S^{(k)}_{\left(c_{1,\delta}, d_{1,\delta}, \ldots, c_{n,\delta}, d_{n,\delta}\right)}\right\}} F(x_1, \cdots, x_n)$

$= \sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1,\cdots,x_n) \in S^{(k)}_{(a_1,b_1,\ldots,a_n,)}\right\}} F(x_1, \cdots, x_n) = \mu_F(R)$.

Tomemos $\delta > 0$ arbitraria, entonces:

$R_\delta \subset [c_{1,\delta}, d_{1,\delta}] \times \cdots \times [c_{n,\delta}, d_{n,\delta}] \subset \bigcup_{i=1}^{\infty} \left(a_1^{(i)}, d_1^{(i)}\right) \times \cdots \times \left(a_n^{(i)}, d_n^{(i)}\right)$.

Así que, por el teorema de Heine-Borel, existe una colección finita, $\left(a_1^{(i_1)}, d_1^{(i_1)}\right) \times \cdots \times \left(a_n^{(i_1)}, d_n^{(i_1)}\right), \ldots, \left(a_1^{(i_m)}, d_1^{(i_m)}\right) \times \cdots \times \left(a_n^{(i_m)}, d_n^{(i_m)}\right)$, tal que:

$R_\delta \subset [c_{1,\delta}, d_{1,\delta}] \times \cdots \times [c_{n,\delta}, d_{n,\delta}] \subset \bigcup_{j=1}^{m} \left(a_1^{(i_j)}, d_1^{(i_j)}\right) \times \cdots \times \left(a_n^{(i_j)}, d_n^{(i_j)}\right)$

$\subset \bigcup_{j=1}^{m} \left(a_1^{(i_j)}, d_1^{(i_j)}\right] \times \cdots \times \left(a_n^{(i_j)}, d_n^{(i_j)}\right] = \bigcup_{j=1}^{m} R_{\delta_{i_j}}$.

Así que:

$\mu_F(R_\delta) \leq \sum_{j=1}^{m} \mu_F\left(R_{\delta_{i_j}}\right) \leq \sum_{i=1}^{\infty} \mu_F(R_{\delta_i})$

$\leq \sum_{i=1}^{\infty} \left[\mu_F\left(R^{(i)}\right) + \frac{\varepsilon}{2^i}\right] = \sum_{i=1}^{\infty} \mu_F\left(R^{(i)}\right) + \varepsilon$.

Y, como $\varepsilon > 0$ es arbitraria:

$\mu_F(R_\delta) \leq \sum_{i=1}^{\infty} \mu_F\left(R^{(i)}\right)$.

Finalmente, tomando límites cuando $\delta \rightsquigarrow 0$, se obtiene:

$\mu_F(R) \leq \sum_{i=1}^{\infty} \mu_F\left(R^{(i)}\right)$.

∎

Teorema 15.2. *Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución finita. Entonces existe una única medida finita μ_F, definida sobre los conjuntos borelianos de $\mathbb{R}^n$, tal que:*

$\mu_F\left((-\infty, x_1] \times \cdots \times (-\infty, x_n]\right) = F(x_1, \ldots, x_n)$

para cualquier $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Demostración

Sea $\mathcal{I}$ la familia de rectángulos de la forma $(a_1, b_1| \times \cdots \times (a_n, b_n|$ y $\mathcal{A}$ la familia de conjuntos de la forma $\bigcup_{j=1}^n R_j$ en donde $n \in \mathbb{N}$ y $R_1, \ldots, R_n$ son rectángulos en $\mathcal{I}$, ajenos por parejas. Para cada rectángulo $R = (a_1, b_1| \times \cdots \times (a_n, b_n| \in \mathcal{I}$, definamos:

$\mu_F\left(R\right) = \sum_{k=0}^n (-1)^k \sum_{\left\{(x_1,\cdots,x_n) \in S^{(k)}_{(a_1,b_1,\ldots,a_n,b_n)}\right\}} F\left(x_1, \cdots, x_n\right)$.

Y, para cada $A = \bigcup_{j=1}^n R_j \in \mathcal{A}$, definamos $\mu_F(A) = \sum_{j=1}^n \mu_F(R_j)$.

Por la proposición 15.2, μ_F está bien definida. Además, $\mathcal{A}$ es un álgebra de subconjuntos de $\mathbb{R}^n$ y la función $\mu_F : \mathcal{A} \mapsto \mathbb{R}$ es no negativa y finitamente aditiva, es decir, es una medida sobre $\mathcal{A}$.

Sea $A_1, A_2, \ldots$ una colección infinita numerable de elementos de $\mathcal{A}$, ajenos por parejas y tales que $A = \bigcup_{i=1}^\infty A_i \in \mathcal{A}$.

Por un lado, como $A = \bigcup_{i=1}^\infty A_i$, A es una unión infinita numerable de rectángulos R_k de la forma $(a_1, b_1| \times \cdots \times (a_n, b_n|$.

Por otro lado, como $A \in \mathcal{A}$, A es una unión finita de rectángulos de la forma $(a_1, b_1| \times \cdots \times (a_n, b_n|$.

Sea $A = \bigcup_{j=1}^n R^{(j)}$, en donde $R^{(j)} = \left(a_1^{(j)}, b_1^{(j)}\right| \times \cdots \times \left(a_n^{(j)}, b_n^{(j)}\right|$.

Para cada $j \in \{1, \ldots, m\}$ y $k \in \mathbb{N}$, definamos $R_k^{(j)} = R_k \cap R^{(j)}$. Entonces, como $\bigcup_{j=1}^n R^{(j)} = \bigcup_{k=1}^\infty R_k$, se tiene $R_k = \bigcup_{j=1}^m R_k^{(j)}$ y $R^{(j)} = \bigcup_{k=1}^\infty R_k^{(j)}$, así que:

$\mu_F\left(A\right) = \sum_{j=1}^m \mu_F\left(R^{(j)}\right) \leq \sum_{j=1}^m \sum_{k=1}^\infty \mu_F\left(R_k^{(j)}\right)$

$= \sum_{k=1}^\infty \sum_{j=1}^m \mu_F\left(R_k^{(j)}\right) = \sum_{k=1}^\infty \mu_F\left(R_k\right) = \sum_{i=1}^\infty \mu_F\left(A_i\right)$.

Además, como μ_F es finitamente aditiva y $A \supset \bigcup_{i=1}^k A_i$ para cualquier $k \in \mathbb{N}$, se tiene $\mu_F(A) \geq \sum_{i=1}^k \mu_F\left(A_i\right)$ para cualquier $k \in \mathbb{N}$, así que:

$\mu_F\left(A\right) \geq \sum_{i=1}^\infty \mu_F\left(A_i\right) T$,

$\mu_F\left(A\right) = \sum_{i=1}^\infty \mu_F\left(A_i\right)$.

Por lo tanto, μ_F es σ-aditiva y entonces puede ser extendida de manera única a una medida μ_F definida sobre la σ-álgebra generada por $\mathcal{A}$, es decir, los borelianos de $\mathbb{R}^n$.

■

Definición 15.4. *Si $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una función distribución finita, la medida μ_F del teorema anterior será llamada la medida generada por F.*

Corolario 15.2. *Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución. Entonces existe un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y una familia $X_1, \ldots, X_n$ de variables aleatorias reales definidas sobre Ω tal que F es la función de distribución conjunta de $X_1, \ldots, X_n$.*

Demostración

Sea $\Omega = \mathbb{R}^n$, $\mathcal{F}$ la σ-álgebra de los conjuntos borelianos de $\mathbb{R}^n$ y P la única medida de probabilidad μ_F, definida sobre los conjuntos borelianos de $\mathbb{R}^n$, tal que

$\mu_F \left((-\infty, x_1] \times \cdots \times (-\infty, x_n] \right) = F(x_1, \ldots, x_n)$

para cualquier $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Para cada $k \in \{1, \ldots, \}$ definamos $X_k : \Omega \mapsto \mathbb{R}$ de la siguiente manera:

$X_k(x_1, \ldots, x_n) = x_k$.

Entonces:

$[X_k \leq x] = (-\infty, x]$.

Así que X_k es una variable aleatoria.

Además:

$P\left[X_1 \leq x_1, \ldots, X_n \leq x_n\right] = \mu_F \left((-\infty, x_1] \times \cdots \times (-\infty, x_n] \right)$

$= F(x_1, \ldots, x_n)$.

Así que $F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = F(x_1, \ldots, x_n)$.

■

Si μ es una medida finita definida sobre los conjuntos borelianos de $\mathbb{R}^n$, la función $F : \mathbb{R}^n \mapsto \mathbb{R}$ definida por:

$F(x_1, \ldots, x_n) = \mu \left((-\infty, x_1] \times \cdots \times (-\infty, x_n] \right)$

es una función de distribución finita y la medida μ_F que genera sobre los borelianos, por ser única, coincide con μ. De esta forma, toda medida finita μ definida sobre los conjuntos borelianos de $\mathbb{R}^n$ está generada por una función de distribución finita en n variables.

15.3. Regularidad de las medidas finitas sobre los borelianos de $\mathbb{R}^n$

PROPOSICIÓN 15.3. *Sea μ una medida finita, definida sobre los conjuntos borelianos de $\mathbb{R}^n$, Entonces, para cualquier rectángulo $R = (a_1, b_1| \times \cdots \times (a_n, b_n|$ en $\mathbb{R}^n$ y $\varepsilon > 0$, existe $\delta > 0$ tal que, si R_δ es el rectángulo $(a_1, d_1) \times \cdots \times (a_n, d_n)$, donde:*

$$d_k = \begin{cases} b_k + \delta & si\ b_k \in \mathbb{R} \\ b_k & si\ b_k = \infty \end{cases}$$

entonces $\mu(R) \leq \mu(R_\delta) < \mu(R) + \varepsilon$

Demostración

Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ la función de distribución finita definida por:

$F(x_1, \ldots, x_n) = \mu((-\infty, x_1] \times \cdots \times (-\infty, x_n])$.

Para cualquier $\delta > 0$ definamos:

$$d_k^{(\delta)} = \begin{cases} b_k + \delta & \text{si } b_k \in \mathbb{R} \\ b_k & \text{si } b_k = \infty \end{cases}$$

$$R^{(\delta)} = \left(a_1, d_1^{(\delta)}\right| \times \cdots \times \left(a_n, d_n^{(\delta)}\right|$$

Se tiene entonces:

$\lim_{\delta \rightsquigarrow 0} \mu\left(R^{(\delta)}\right)$

$= \lim_{\delta_i \rightsquigarrow 0} \sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1, \cdots, x_n) \in S^{(k)}_{\left(a_1, d_1^\delta, \ldots, a_n, d_n^\delta\right)}\right\}} F(x_1, \cdots, x_n)$

$= \sum_{k=0}^{n} (-1)^k \sum_{\left\{(x_1, \cdots, x_n) \in S^{(k)}_{(a_1, b_1, \ldots, a_n, b_n)}\right\}} F(x_1, \cdots, x_n) = \mu(R)$.

Dada $\varepsilon > 0$, existe entonces $\delta > 0$ tal que:

$\mu\left(R^{(\delta)}\right) - \mu(R) < \varepsilon$.

Finalmente, como $R \subset R_\delta \subset R^{(\delta)}$, se tiene:

$\mu(R) \leq \mu(R_\delta) \leq \mu\left(R^{(\delta)}\right) < \mu(R) + \varepsilon$.

■

TEOREMA 15.3. *Sea μ una medida finita, definida sobre los conjuntos borelianos de $\mathbb{R}^n$, Entonces, para cualquier conjunto boreliano B de $\mathbb{R}^n$ se tiene:*

$\mu(B) = \inf\{\mu(O) : B \subset O$ *y O es un abierto de $\mathbb{R}^n$*$\}$.

Demostración

Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ la función de distribución finita definida por:

$F(x_1, \ldots, x_n) = \mu\left((-\infty, x_1] \times \cdots \times (-\infty, x_n]\right)$.

Sea $\mathcal{I}$ la familia de rectángulos de la forma $(a_1, b_1| \times \cdots \times (a_n, b_n|$ y $\mathcal{A}$ la familia de conjuntos de la forma $\bigcup_{j=1}^{n} R_j$ en donde $n \in \mathbb{N}$ y $R_1, \ldots, R_n$ son rectángulos en $\mathcal{I}$, ajenos por parejas.

$\mathcal{A}$ es un álgebra de subconjuntos de $\mathbb{R}^n$ y la σ-álgebra generada por $\mathcal{A}$ es la σ-álgebra de conjuntos borelianos en $\mathbb{R}^n$.

Como μ es la medida generada por F, se tiene:

$\mu(B) = \text{ínf}\left\{\sum_i \mu(A_i) : A_1, A_2, \ldots \in \mathcal{A} \text{ y } B \subset \bigcup_i A_i\right\}$.

para cualquier conjunto boreliano B de $\mathbb{R}^n$.

Sea B un conjunto boreliano de $\mathbb{R}^n$. Dada $\varepsilon > 0$, consideremos una colección, $A_1, A_2, \ldots$, de elementos de $\mathcal{A}$ tal que:

$B \subset \bigcup_i A_i$,

$\sum_i \mu(A_i) < \mu(B) + \frac{\varepsilon}{2}$.

Cada A_i es de la forma:

$A_i = \bigcup_{j=1}^{n_i} R^{(i,j)}$,

donde $n_i \in \mathbb{N}$ y $R^{(i,1)}, \ldots, R^{(i,n_i)}$ son rectángulos en $\mathcal{I}$ ajenos por parejas.

Para cada rectángulo $R^{(i,j)} = \left(a_1^{(i,j)}, b_1^{(i,j)}\right| \times \cdots \times \left(a_n^{(i,j)}, b_n^{(i,j)}\right|$ sea $\delta_{ij} > 0$ tal que, si $R_{\delta_{ij}}^{(i,j)}$ es el rectángulo $\left(a_1^{(i,j)}, d_1^{(i,j)}\right) \times \cdots \times \left(a_n^{(i,j)}, d_n^{(i,j)}\right)$, donde:

$$d_k^{(i,j)} = \begin{cases} b_k^{(i,j)} + \delta_{ij} & \text{si } b_k^{(i,j)} \in \mathbb{R} \\ b_k & \text{si } b_k^{(i,j)} = \infty \end{cases}$$

Entonces $\mu\left(R_{\delta_{ij}}^{(i,j)}\right) < \mu\left(R^{(i,j)}\right) + \frac{\varepsilon}{2^{i+j+1}}$.

Definamos:

$O_\varepsilon = \bigcup_i \bigcup_{j=1}^{n_i} R_{\delta_{ij}}^{(i,j)}$.

Entonces O_ε es un abierto de $\mathbb{R}^n$ que contiene a B y se tiene:

$\mu\left(O_\varepsilon\right) \leq \sum_i \sum_{j=1}^{n_i} \mu\left(R_{\delta_{ij}}^{(i,j)}\right) < \sum_i \left(\sum_{j=1}^{n_i} \mu\left(R^{(i,j)}\right) + \sum_{j=1}^{n_i} \frac{\varepsilon}{2^{i+j+1}}\right)$

$\leq \sum_i \left(\mu(A_i) + \frac{\varepsilon}{2^{i+1}}\right) \leq \sum_i \mu(A_i) + \frac{\varepsilon}{2}$

$< \mu(B) + \varepsilon.$

■

TEOREMA 15.4. *Sea μ una medida finita, definida sobre los conjuntos borelianos de $\mathbb{R}^n$, Entonces, para cualquier conjunto boreliano B de $\mathbb{R}^n$ se tiene:*

$\mu(B) = \sup\{\mu(K) : K \subset B \text{ y } K \text{ es un compacto de } \mathbb{R}^n\}.$

Demostración

Sean B un boreliano de $\mathbb{R}^n$ y $K_1, K_2, \ldots$ una sucesión creciente de conjuntos compactos de $\mathbb{R}^n$ tal que $\bigcup_{i=1}^{\infty} K_i = \mathbb{R}^n$.

Dada $\varepsilon > 0$, para cada $i \in \mathbb{N}$, sea O_i un abierto de $\mathbb{R}^n$ tal que $K_i \cap B^c \subset O_i$ y $\mu(O_i) < \mu(K_i \cap B^c) + \frac{\varepsilon}{2}$. Entonces $K_i \cap O_i^c$ es un compacto de $\mathbb{R}^n$ y se tiene:

$K_i \cap O_i^c \subset K_i \cap (K_i \cap B^c)^c = K_i \cap (K_i^c \cup B) = K_i \cap B.$

$K_i \cap B - K_i \cap O_i^c = (K_i \cap B) \cap (K_i \cap O_i^c)^c$

$= (K_i \cap B) \cap (K_i^c \cup O_i) = O_i \cap (K_i \cap B)$

$= (O_i - K_i \cap B^c) \cap (K_i \cap B) \subset O_i - K_i \cap B^c.$

Así que:

$\mu(K_i \cap B) - \mu(K_i \cap O_i^c) \leq \mu(O_i) - \mu(K_i \cap B^c) < \frac{\varepsilon}{2}.$

Por otra parte, $B = \bigcup_{i=1}^{\infty} (K_i \cap B)$, así que:

$\mu(B) = \sup_{i \in \mathbb{N}} \mu(K_i \cap B).$

Sea $N \in \mathbb{N}$ tal que $\mu(B) < \mu(K_N \cap B) + \frac{\varepsilon}{2}$, entonces $K_N \cap O_N^c$ es un compacto contenido en B y se tiene:

$\mu(B) < \mu(K_N \cap B) + \frac{\varepsilon}{2} < \mu(K_N \cap O_N^c) + \varepsilon.$

■

15.4. Sucesiones de variables aleatorias independientes

Para formular los otros dos ejemplos dentro del marco teórico expuesto en el capítulo anterior, requerimos construir un espacio de probabilidad $(\Omega, \Im, P)$ en el cual se pueda definir una infinidad numerable de variables aleatorias independientes, cada una con distribución Bernoulli de parámetro p. Vamos a mostrar esta construcción para el caso $p = \frac{1}{2}$. En el caso general la construcción es similar, pero requiere de algunos pasos adicionales, Primero un resultado que usaremos más adelante:

Teorema 15.5. *Sea* $(\Omega, \mathcal{A}, P)$ *un espacio de probabilidad y* $X_1, X_2, \ldots$ *una sucesión de variables aleatorias independientes, definidas sobre ese espacio, cada una de ellas con distribución Bernoulli de parámetro* $\frac{1}{2}$. *Definamos* $X : \Omega \mapsto \mathbb{R}$ *mediante la relación* $X = \sum_{k=1}^{\infty} \frac{X_k}{2^k}$. *Entonces,* X *es una variable aleatoria con distribución uniforme en el intervalo* $[0,1]$.

Demostración

X es el límite de la sucesión no decreciente de variables aleatorias $X_n = \sum_{k=1}^{n} \frac{X_k}{2^k}$, así que es ella misma una variable aleatoria.

Obsérvese que si vemos cada sucesión $(X_k(\omega))$ como el desarrollo en base 2 de un número real en el intervalo $[0,1]$, $X(\omega)$ es precisamente ese número real.

Si $(s_n)_{n\in\mathbb{N}}$ es una sucesión de 0's y 1's, se tiene $P\left[X_k = s_k \text{ para toda } k \in \mathbb{N}\right] = 0$ y si $x \in (0,1]$ es un racional diádico, x tiene exactamente dos desarrollos en base 2, así que $P\left[X \in x\right] = 0$. Además, $x = 0$ tiene únicamente un desarrollo en base 2, así que, también, $P\left[X = 0\right] = 0$. Por lo tanto, $P\left[X \in x\right] = 0$ para cualquier racional diádico $x \in [0,1]$.

Por otra parte, para cada $n \in \mathbb{N}$, consideremos un intervalo de la forma $\left(\frac{j-1}{2^n}, \frac{j}{2^n}\right)$, con $j \in \{1, \ldots, 2^n\}$. Asociada a tal intervalo existe una única colección de 0's y 1's, $(s_1, s_2, \ldots, s_n)$, tal que un punto x pertenece al intervalo $\left(\frac{j-1}{2^n}, \frac{j}{2^n}\right)$ si y sólo si tiene un desarrollo en base 2 de la forma $x = 0.s_1 s_2 \ldots s_n \ldots$. Por lo tanto:

$$P\left[X \in \left(\tfrac{j-1}{2^n}, \tfrac{j}{2^n}\right)\right] = P\left[X_1 = s_1, X_2 = s_2, \ldots, X_n = s_n\right] = \tfrac{1}{2^n}.$$

Así que, si $k \in \{1, \ldots, 2^n\}$, se tiene:

$$P\left[X \in \left(0, \tfrac{k}{2^n}\right)\right] = \sum_{j=1}^{k} P\left[X \in \left(\tfrac{j-1}{2^n}, \tfrac{j}{2^n}\right)\right] = \sum_{j=1}^{k} \tfrac{1}{2^n} = \tfrac{k}{2^n}.$$

Es decir, si $x \in (0,1]$ es un racional diádico, se tiene:

$$P\left[X \in (0,x)\right] = x.$$

Combinando este resultado con el anterior, se tiene:

$$P\left[X \in (0,x]\right] = x$$

para cualquier racional diádico $x \in (0,1]$.

Además, la función $F_X : [0,1] \to \mathbb{R}$, definida por:

$$F_X(x) = P\left[X \in (0,x]\right],$$

es continua por la derecha.

Por lo tanto:

$F_X(x) = x$ para cualquier $x \in [0,1]$.

Así que X tiene distribución uniforme en el intervalo $[0,1]$.

■

En lo que sigue vamos a tomar como espacio de probabilidad a la terna $([0,1], \mathcal{L}, P)$, donde P es la medida de Lebesgue en el intervalo $[0,1]$.

Para cada $n \in \mathbb{N}$, definamos:

$\mathbb{B}_n = \{(s_1, s_2, \ldots, s_n) : s_j \in \{0,1\} \text{ para cualquier } j \in \{1, 2, \ldots, n\}\}$, donde $n \in \mathbb{N}$.

Sabemos que cada $x \in [0,1]$ tiene un desarrollo en base 2, es decir, se puede expresar de la siguiente manera:

$x = \sum_{k=1}^{\infty} \frac{s_k}{2^k}$,

donde $s_k \in \{0,1\}$ para cualquier $k \in \mathbb{N}$.

Para los números reales $x \in (0,1)$ de la forma $x = \frac{j}{2^n}$, donde $n \in \mathbb{N}$ y $j \in \{1, 2, \ldots, 2^n - 1\}$, es decir, para los racionales diádicos, el desarrollo en base 2 no es único. Para cada uno de estos puntos, elijamos como desarrollo en base 2 a la sucesión $s_1, s_2, \ldots$ para la cual existe $N \in \mathbb{N}$ tal que $s_k = 1$ para cualquier $k \in \{N+1, N+2, \ldots\}$.

Para $x = 0$, el desarrollo en base 2 es la sucesión $(s_n)_{n\in\mathbb{N}}$ idénticamente cero, mientras que para $x = 1$ es la sucesión $(s_n)_{n\in\mathbb{N}}$ idénticamente uno.

De esta forma, el desarrollo en base 2 de un número real $x \in [0,1]$ es único y, si $x \in (0,1]$, consiste de una sucesión $(s_n)_{n\in\mathbb{N}}$ tal que $s_k = 1$ para una infinidad de índices k.

Obsérvese que si $(s_1, s_2, \ldots, s_n) \in \mathbb{B}_n$ y $x \in (0,1]$, entonces, $s_1, s_2, \ldots, s_n$ son los primeros n términos del desarrollo de x en base 2 si y sólo si $x \in \left(\sum_{i=1}^{n} \frac{s_i}{2^i}, \sum_{i=1}^{n} \frac{s_i}{2^i} + \frac{1}{2^n}\right]$. Además cada uno de los 2^n intervalos de la forma $\left(\sum_{i=1}^{n} \frac{s_i}{2^i}, \sum_{i=1}^{n} \frac{s_i}{2^i} + \frac{1}{2^n}\right]$, donde $(s_1, s_2, \ldots, s_n) \in \mathbb{B}_n$, es alguno de los intervalos de la forma $\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]$, donde $k \in \{1, 2, \ldots, 2^n\}$, los cuales constituyen una partición del intervalo $(0,1]$.

Teorema 15.6. *Para cada $n \in \mathbb{N}$ definamos $X_n : [0,1] \to \mathbb{R}$ mediante la relación $X_n(x) = s_n$, donde s_n es el n-simo término del desarrollo en base 2 de x. Entonces, las variables aleatorias de la familia $\{X_n : n \in \mathbb{N}\}$ son independientes y cada una de ellas tiene distribución Bernoulli con parámetro $p = \frac{1}{2}$.*

Demostración

Si $r \in \{0,1\}$, se tiene:

$P[X_n = r] = \sum_{\{(r_1, r_2, \ldots, r_{n-1}) \in \mathbb{B}_{n-1}\}} P[X_1 = r_1, X_2 = r_2, \ldots, X_{n-1} = r_{n-1}, X_n = r]$

$= \sum_{\{(r_1, r_2, \ldots, r_{n-1}) \in \mathbb{B}_{n-1}\}} P\left(\left(\sum_{k=1}^{n-1} \frac{r_k}{2^k} + \frac{r}{2^n}, \sum_{k=1}^{n-1} \frac{r_k}{2^k} + \frac{r}{2^n} + \frac{1}{2^n}\right)\right)$

$= \sum_{\{(r_1,r_2,\ldots,r_{n-1})\in\mathbb{B}_{n-1}\}} \frac{1}{2^n} = \frac{1}{2}$.

Por lo tanto, para cualquier $n \in \mathbb{N}$, X_n tiene distribución Bernoulli con parámetro $p = \frac{1}{2}$.

Además, para cualquier $m \in \mathbb{N}$ y $(r_1, r_{2,} \ldots, r_m) \in \mathbb{B}_m$, se tiene:

$P\left(\bigcap_{j=1}^m [X_j = r_j]\right) = P\left(\left(\sum_{k=1}^m \frac{r_k}{2^k}, \sum_{k=1}^m \frac{r_k}{2^k} + \frac{1}{2^m}\right)\right) = \frac{1}{2^m} = \prod_{j=1}^m P[X_j = r_j]$.

Así que las variables aleatorias $X_1, X_2, \ldots, X_m$ son independientes.

Por lo tanto, las variables aleatorias de la familia $\{X_n : n \in \mathbb{N}\}$ son independientes y cada una de ellas tiene distribución Bernoulli con parámetro $p = \frac{1}{2}$.

■

TEOREMA 15.7. *Se puede definir, sobre* $([0,1], \mathcal{L}, P)$*, una sucesión de variables aleatorias independientes, cada una de ellas con distribución uniforme en el intervalo* $(0,1)$*.*

Demostración

Consideremos una sucesión $X_1, X_2, \ldots$ de variables aleatorias independientes, definidas sobre $([0,1], \mathcal{L}, P)$, cada una de ellas con distribución Bernoulli de parámetro $\frac{1}{2}$.

De acuerdo con la primera proposición, si $(X_{n_k})_{k\in\mathbb{N}}$ es cualquier subsucesión de la sucesión $(X_n)_{n\in\mathbb{N}}$ y definimos $Y : [0,1] \to \mathbb{R}$ mediante la relación $Y = \sum_{k=1}^\infty \frac{Y_k}{2^k}$, entonces Y es una variable aleatoria con distribución uniforme en el intervalo $(0,1)$.

Para definir una sucesión de variables aleatorias independientes, definidas sobre $([0,1], \mathcal{L}, P)$, cada una de ellas con distribución uniforme en el intervalo $(0,1)$, basta con mostrar que existe una infinidad numerable de subsucesiones de $(X_n)_{n\in\mathbb{N}}$ de tal manera que cualquier par de ellas no tengan elementos en común. Una vez mostrado esto, cada una de esas subsucesiones genera una distribución uniforme en el intervalo $(0,1)$.

Existen diferentes maneras de tomar las subsucesiones de $(X_n)_{n\in\mathbb{N}}$ con la propiedad mencionada. Por ejemplo, si $\{p_1, p_2, \ldots\}$ es el conjunto de números primos mayores que 1 y, para cada $n \in \mathbb{N}$, definimos $A_n = \{p_n, p_n^2, p_n^3, \ldots\}$, entonces los conjuntos A_n son ajenos por parejas, así que las subsucesiones $\left(X_{p_1^k}\right)_{k\in\mathbb{N}}, \left(X_{p_2^k}\right)_{k\in\mathbb{N}}, \left(X_{p_3^k}\right)_{k\in\mathbb{N}}, \ldots$ cumplen con la propiedad requerida.

También podemos ordenar los elementos de la sucesión $(X_n)_{n\in\mathbb{N}}$ de la siguiente manera:

$X_1, X_2, X_4, X_7, X_{11}, X_{16}, X_{22}, \ldots$

$X_3, X_5, X_8, X_{12}, X_{17}, X_{23}, \ldots$

$X_6, X_9, X_{13}, X_{18}, X_{24}, \ldots$

$X_{10}, X_{14}, X_{19}, X_{25}, \ldots$

$X_{15}, X_{20}, X_{26}, \ldots$

$X_{21}, X_{27}, \ldots$

$X_{28}, \ldots$

$\vdots$

Cada renglón forma una subsucesión de $(X_n)_{n\in\mathbb{N}}$ y las subsucesiones de dos renglones diferentes no tienen elementos en común.

El n-simo renglón está dado por:

$X_{\frac{1}{2}n(n+1)}, X_{\frac{1}{2}n(n+1)+n}, X_{\frac{1}{2}n(n+1)+n+(n+1)}, X_{\frac{1}{2}n(n+1)+n+(n+1)+(n+2)},$

$\ldots, X_{\frac{1}{2}n(n+1)+n+(n+1)+(n+2)+\cdots+(n+k-1)}, \ldots.$

Es decir, el k-ésimo elemento del n-simo renglón está dado por:

$X_{\frac{1}{2}n(n+1)+\sum_{j=1}^{k-1}(n+j-1)} = X_{\frac{1}{2}n(n+1)+n(k-1)+\frac{1}{2}(k-1)(k-2)}.$

Así que, las sucesiones $\left(X_{1+(k-1)+\frac{1}{2}(k-1)(k-2)}\right)_{k\in\mathbb{N}}, \left(X_{3+2(k-1)+\frac{1}{2}(k-1)(k-2)}\right)_{k\in\mathbb{N}},$

$\left(X_{6+3(k-1)+\frac{1}{2}(k-1)(k-2)}\right)_{k\in\mathbb{N}}, \ldots$ cumplen con las propiedades requeridas.

De manera general, si las sucesiones $\left(X_k^{(n)}\right)_{k\in\mathbb{N}}, \left(X_k^{(2)}\right)_{k\in\mathbb{N}}, \left(X_k^{(3)}\right)_{k\in\mathbb{N}}, \ldots$ son subsucesiones de $(X_n)_{n\in\mathbb{N}}$ tales que cualquier par de ellas no tienen elementos en común, definamos, para cada $n \in \mathbb{N}$:

$U_n = \sum_{k=1}^{\infty} \frac{X_k^{(n)}}{2^k}.$

Entonces, para cada $n \in \mathbb{N}$, U_n tiene distribución uniforme en el intervalo $(0,1)$.

Para cada $m \in \mathbb{N}$, las sumas parciales $\sum_{k=1}^{m} \frac{X_k^{(1)}}{2^k}, \sum_{k=1}^{m} \frac{X_k^{(2)}}{2^k}, \sum_{k=1}^{m} \frac{X_k^{(3)}}{2^k}, \ldots$ forman una familia de variables aleatorias independientes, así que, para cualquier $n \in \mathbb{N}$ y cualquier $(x_1, \ldots, x_n) \in \mathbb{R}^n$, se tiene:

$P\left[\sum_{k=1}^{m} \frac{X_k^{(1)}}{2^k} \leq x_1, \ldots, \sum_{k=1}^{m} \frac{X_k^{(n)}}{2^k} \leq x_n\right] = P\left[\sum_{k=1}^{m} \frac{X_k^{(1)}}{2^k} \leq x_1\right] \cdots P\left[\sum_{k=1}^{m} \frac{X_k^{(n)}}{2^k} \leq x_n\right].$

Como, para cada $j \in \mathbb{N}$, la sucesión $\left(\sum_{k=1}^{m} \frac{X_k^{(j)}}{2^k}\right)_{m\in\mathbb{N}}$ es no decreciente, la sucesión de eventos $\left(\left[\sum_{k=1}^{m} \frac{X_k^{(1)}}{2^k} \leq x_1, \ldots, \sum_{k=1}^{m} \frac{X_k^{(n)}}{2^k} \leq x_n\right]\right)_{m\in\mathbb{N}}$ es decreciente y la intersección de todos ellos es el evento $\left[\sum_{k=1}^{\infty} \frac{X_k^{(1)}}{2^k} \leq x_1, \ldots, \sum_{k=1}^{\infty} \frac{X_k^{(n)}}{2^k} \leq x_n\right]$; así que:

$$P\left[\sum_{k=1}^{\infty} \frac{X_k^{(1)}}{2^k} \leq x_1, \ldots, \sum_{k=1}^{\infty} \frac{X_k^{(n)}}{2^k} \leq x_n\right] = \text{lím}_{m\rightsquigarrow\infty} P\left[\sum_{k=1}^{m} \frac{X_k^{(1)}}{2^k} \leq x_1, \ldots, \sum_{k=1}^{m} \frac{X_k^{(n)}}{2^k} \leq x_n\right]$$

$$= \left(\text{lím}_{m\rightsquigarrow\infty} P\left[\sum_{k=1}^{m} \frac{X_k^{(1)}}{2^k} \leq x_1\right]\right) \cdots \left(\text{lím}_{m\rightsquigarrow\infty} P\left[\sum_{k=1}^{m} \frac{X_k^{(n)}}{2^k} \leq x_n\right]\right)$$

$$= P\left[\sum_{k=1}^{\infty} \frac{X_k^{(1)}}{2^k} \leq x_1\right] \cdots P\left[\sum_{k=1}^{\infty} \frac{X_k^{(n)}}{2^k} \leq x_n\right].$$

Así que las variables aleatorias $U_1, U_2, \ldots$ son independientes.

∎

TEOREMA 15.8. *Sea $(\Omega, \mathcal{L}, \lambda)$ el espacio de probabilidad formado por $\Omega = (0,1)$, $\mathcal{L}$ la σ-álgebra de los conjuntos Lebesgue medibles en el intervalo $(0,1)$ y λ la medida de Lebesgue el intervalo $(0,1)$. Sea, además, $(\mu_n)_{n\in\mathbb{N}}$ una sucesión de medidas de probabilidad sobre $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. Existe entonces una sucesión de variables aleatorias reales independientes $(X_n)_{n\in\mathbb{N}}$, definidas sobre $(\Omega, \mathcal{L}, \lambda)$, tales que, para cualquier $n \in \mathbb{N}$, la distribución de X_n es μ_n.*

Demostración

Consideremos una sucesión de variables aleatorias independientes $(U_n)_{n\in\mathbb{N}}$, definidas sobre $(\Omega, \mathcal{B}, \lambda)$, cada una de ellas con distribución uniforme en el intervalo $(0,1)$.

Para cada $n \in \mathbb{N}$, definamos las funciones $F_n : \mathbb{R} \mapsto \mathbb{R}$ y $c_n : (0,1) \mapsto \mathbb{R}$ mediante las siguientes relaciones;

$F_n(x) = \mu_n((-\infty, x])$,

$c_n(t) = \text{ínf}\{x \in \mathbb{R} : F_n(x) > t\}$.

Para cada $n \in \mathbb{N}$, definamos $X_n = c_n(U_n)$.

Por la proposición ...., la función de distribución de X_n es F_n. Además, μ_n es la medida generada por F_n, así que la distribución de X_n es μ_n.

Finalmente, como las variables aleatorias $U_1, U_2, \ldots$ son independientes, también lo son $X_1, X_2, \ldots$.

∎

15.5. Sucesiones de variables aleatorias con distribuciones finito dimensionales conocidas

Sea $(\Omega, \mathcal{B}, \lambda)$ el espacio de probabilidad formado por $\Omega = (0,1)$, $\mathcal{B}$ la σ-álgebra de los conjuntos borelianos del intervalo $(0,1)$ y λ la medida de Lebesgue el intervalo $(0,1)$. Sea $U_1, U_2, \ldots$ una sucesión de variables aleatorias independientes, definidas sobre ese espacio, cada una de ellas con distribución uniforme en el intervalo $(0,1)$.

Sea $T = \{t_1, t_2, \ldots\}$ un conjunto infinito numerable y supongamos que para cada subconjunto finito de T, $u = \{u_1, \ldots, u_n\}$, se tiene una función de distribución conjunta $F_u : \mathbb{R}^n \to \mathbb{R}$ de tal forma que si $u = \{u_1, \ldots, u_n\}$ y $v = \{v_1, \ldots, v_m\}$ son dos subconjuntos finitos de T tales que $u \subset v$, entonces la función de distribución conjunta F_u coincide con la distribución conjunta marginal que se obtiene de F_v restringiéndola a los elementos de u.

Para cada $n \in \mathbb{N}$, denotemos por $\mathcal{B}_n$ a la σ-álgebra de Borel en $\mathbb{R}^n$, por F_n a la función de distribución conjunta $F_{\{t_1,\ldots,t_n\}}$ y por μ_n a la medida, definida sobre $(\mathbb{R}^n, \mathcal{B}_n)$, generada por F_n.

Vamos a demostrar que existe una sucesión de variables aleatorias $X_1, X_2, \ldots$, definidas sobre $(\Omega, \mathcal{B}, \lambda)$, tal que, para cada $n \in \mathbb{N}$, la función de distribución conjunta del vector aleatorio $(X_1, \ldots, X_n)$ es F_n y existe una función medible $d_n : (\mathbb{R}^n, \mathcal{B}_n) \mapsto \mathbb{R}$ tal que $X_n = d_n(U_n, X_1, \ldots, X_{n-1})$.

Definamos:

$d_1(t) = \text{ínf}\{s \in \mathbb{R} : F_1(s) \geq t\}$,

$X_1 = d_1(U_1)$.

Sabemos que la función de distribución de X_1 es F_1.

Supongamos que tenemos definidas, sobre $(\Omega, \mathcal{B}, \lambda)$, $n-1$ variables aleatorias $X_1, \ldots, X_{n-1}$ cuya función de distribución conjunta es F_{n-1} y tales que, para cada $k \in \{1, \ldots, n-1\}$, existe una función medible $d_k : \left(\mathbb{R}^k, \mathcal{B}_k\right) \mapsto \mathbb{R}$ tal que $X_k = d_k(U_k, X_1, \ldots, X_{k-1})$.

Para cada $k \in \{1, \ldots, n\}$, sea Y_k la proyección de $\mathbb{R}^n$ sobre la k-ésima coordenada.

La función de distribución conjunta de $Y_1, \ldots, Y_n$ es F_n.

Definamos:

$G_n(Y_1, \ldots, Y_{n-1}, y) = \mu_n[Y_n \leq y \mid Y_1, \ldots, Y_{n-1}]$.

Si $A \in \mathcal{B}_{n-1}$, se tiene:

$P[(Y_1, \ldots, Y_{n-1}) \in A, Y_n \leq y] = E[I_A(Y_1, \ldots, Y_{n-1})\, G_n(Y_1, \ldots, Y_{n-1}, y)]$

$= \int_A G_n(y_1, \ldots, y_{n-1}, y)\, dF_{n-1}(y_1, \ldots, y_{n-1})$.

Definamos:

$d_n(t, y_1, \dots, y_{n-1}) = \text{ínf}\,\{s \in \mathbb{R} : F_n\,(s, y_1, \dots, y_{n-1}) \geq t\}$

$X_n = d_n(U_n, X_1, \dots, X_{n-1})$.

Si $A \in \mathcal{B}_{n-1}$, se tiene:

$P\,[(X_1, \dots, X_{n-1}) \in A, X_n \leq x] = E\,[I_A\,(X_1, \dots, X_{n-1})\,, d_n(U_n, X_1, \dots, X_{n-1}) \leq x]$

$= E\,[I_A\,(X_1, \dots, X_{n-1})\,, U_n \leq G_n\,(X_1, \dots, X_{n-1}, x)]$

$= \int_A \int_0^{G_n(x_1,\dots,x_{n-1},x)} dF_U(u) dF_{n-1}(x_1, \dots, x_{n-1})$

$= \int_A G_n\,(x_1, \dots, x_{n-1}, x)\, dF_{n-1}(x_1, \dots, x_{n-1})$

$= P\,[(Y_1, \dots, Y_{n-1}) \in A, Y_n \leq x]$.

Así que la distribución del vector aleatorio $(X_1, \dots, X_n)$ es la misma que la del vector aleatorio $(Y_1, \dots, Y_n)$.

15.6. Teorema de Kolmogorov

En esta sección demostraremos el teorema de Kolmogorov, el cual asegura la existencia de un espacio de probabilidad asociado a una familia cualquiera de variables aleatorias con distribuciones finito dimensionales conocidas.

La idea de la demostración es la siguiente:

Para cada subconjunto finito, de un conjunto infinito Γ, se tiene una función de distribución finito dimensional, de tal manera que se satisface la condición de consistencia que se formula en el enunciado del teorema. Se considera entonces el producto cartesiano de tantas copias de $\mathbb{R}$ como elementos tenga Γ, es decir $\mathbb{R}^\Gamma$. Para cada subconjunto finito u de Γ se expresa $\mathbb{R}^\Gamma$ como el producto cartesiano $\mathbb{R}^u \times \mathbb{R}^{\Gamma - u}$ y se genera, sobre $\mathbb{R}^u$, una medida de probabilidad a partir de la distribución finito dimensional correspondiente al conjunto u (es decir, se genera una medida sobre los borelianos de $\mathbb{R}^u$). De lo que se trata entonces es de obtener una medida sobre $\mathbb{R}^\Gamma$ juntando todas las medidas que se obtienen sobre los conjuntos $\mathbb{R}^u$, donde u corre sobre todos los subconjuntos finitos de Γ. Esto se logra definiendo primero una cuasi medida sobre el álgebra de subconjuntos de $\mathbb{R}^\Gamma$ formada por la familia de todos los conjuntos que pertenecen a $\mathfrak{B}\,(\mathbb{R}^u) \times \mathbb{R}^{\Gamma - u}$ para algún subconjunto finito u de Γ. Después se aplica el teorema de extensión de Carathéodory para obtener una medida sobre la σ-álgebra generada por esa álgebra.

TEOREMA 15.9 (**Teorema de Kolmogorov**). *Sea Γun conjunto infinito y supongamos que para cada subconjunto finito de Γ, $u = \{t_1, \dots, t_n\}$, se tiene una función de distribución conjunta $F_u : \mathbb{R}^n \to \mathbb{R}$ de tal forma que si $u = \{t_1, \dots, t_n\}$ y $v = \{s_1, \dots, s_m\}$ son dos subconjuntos finitos de T tales que $u \subset v$, entonces la función de distribución conjunta F_u*

coincide con la distribución conjunta marginal que se obtiene de F_v restringiéndola a las coordenadas en u. Entonces, existe un espacio de probabilidad $(\Omega, \Im, P)$ y una familia de variables aleatorias reales $\{X_t\}_{t\in T}$ definidas sobre Ω tal que si $u = \{t_1, \dots, t_n\}$ es cualquier subconjunto finito de Γ, entonces la función de distribución conjunta de $X_{t_1}, \dots, X_{t_n}$ es F_u.

Demostración

Denotemos por U a la familia de subconjuntos finitos de Γ.

Sean $\Omega = \mathbb{R}^\Gamma = \{f : \Gamma \mapsto \mathbb{R}\}$y, para cada $u \in U$, $\mathbb{R}^u = \{f : u \mapsto \mathbb{R}\}$.

Por definición, un elemento $\omega \in \Omega$ es una función de Γ en $\mathbb{R}$, sin embargo podemos también imaginar a ω como un vector el cual tiene una coordenada para cada $t \in \Gamma$.

Por la proposición 15.2, sabemos que, para cada $u \in U$, existe una única medida de probabilidad P_u definida sobre los conjuntos borelianos de $\mathbb{R}^u$ tal que:

$P_u\left((-\infty, x_1] \times \cdots \times (-\infty, x_n]\right) = F_u(x_1, \dots, x_n)$

para cualquier $(x_1, \dots, x_n) \in \mathbb{R}^u$.

Para cada $u = \{u_1, \dots, u_n\} \in U$, denotemos por Π_u a la función $\Pi_u : \Omega \mapsto \mathbb{R}^u$ definida por $\Pi_u(\omega) = (\omega(u_1), \dots, \omega(u_n))$ y, para pareja $u, v \in U$ tal que $u \subset v$, denotemos por Π_{vu} a la función $\Pi_{vu} : \mathbb{R}^v \mapsto \mathbb{R}^u$ definida por $\Pi_{vu}(f) = f_u$, donde $f_u : u \mapsto \mathbb{R}$ es la restricción de $f : v \mapsto \mathbb{R}$ a u. Obsérvese que Π_u y Π_{vu} son simplemente proyecciones sobre un espacio de menos coordenadas al del dominio.

Sabemos que si $u, v \in U$ y $u \subset v$, entonces la función de distribución conjunta F_u coincide con la distribución conjunta marginal que se obtiene de F_v restringiéndola a las coordenadas de u. Esto se traduce en la relación $P_u(B_u) = P_v\left(\Pi_{vu}^{-1}(B_u)\right)$ para cualquier conjunto boreliano B_u de $\mathbb{R}^u$.

Para cada $u \in U$, denotemos por $\mathcal{B}_u$ a la σ-álgebra de los conjuntos borelianos de $\mathbb{R}^u$ y definamos:

$\Im_0 = \{\Pi_u^{-1}(B_u) : u \in U \text{ y } B_u \in \mathcal{B}_u\}$.

Cada elemento de $\Im_0$ es un subconjunto de Ω, es decir está formado por vectores cada uno de los cuales tiene una coordenada para cada $t \in \Gamma$; lo que caracteriza a esos vectores es que restringiéndonos a las coordenadas que corresponden a los elementos de u, se obtiene un elemento de B_u. También puede pensarse $\Pi_u^{-1}(B_u)$ como $B_u \times \mathbb{R}^{\Gamma - u}$; es decir, restringiéndonos a las coordenadas correspondientes a u, $\Pi_u^{-1}(B_u)$ es B_u, mientras que restringiéndonos a las coordenadas correspondientes a $\Gamma - u$ es $\mathbb{R}^{\Gamma - u}$.

Obviamente, $\Omega \in \Im_0$ y si $E \in \Im_0$ entonces $E^c \in \Im_0$. Por otra parte, si $E = \Pi_u^{-1}(A_u) \in \Im_0$ y $F = \Pi_v^{-1}(B_v) \in \Im_0$, sea $w = u \cup v$, $A_w = \Pi_{wu}^{-1}(A_u)$ y $B_w = \Pi_{wv}^{-1}(B_v)$, entonces:

$E \cup F = \Pi_w^{-1}(A_w \cup B_w) \in \Im_0$.

Por lo tanto, $\Im_0$ es un álgebra de subconjuntos de Ω.

Definamos $P : \Im_0 \mapsto [0,1]$ de la siguiente manera:

$P\left(\Pi_u^{-1}\left(B_u\right)\right) = P_u\left(B_u\right)$.

Observemos en primer lugar que P está bien definida. En efecto, supongamos que $\Pi_u^{-1}\left(B_u\right) = \Pi_v^{-1}\left(A_v\right)$, entonces, definiendo $w = u \cup v$, se tiene $\Pi_u^{-1}\left(B_u\right) = \Pi_w^{-1}\left(\Pi_{wu}^{-1}\left(B_u\right)\right)$ y $\Pi_v^{-1}\left(A_v\right) = \Pi_w^{-1}\left(\Pi_{wv}^{-1}\left(A_v\right)\right)$, así que $\Pi_{wu}^{-1}\left(B_u\right) = \Pi_{wv}^{-1}\left(A_v\right)$. Por lo tanto:

$P_u(B_u) = P_w\left(\Pi_{wu}^{-1}\left(B_u\right)\right) = P_w\left(\Pi_{wv}^{-1}\left(A_v\right)\right) = P_v(A_v)$.

Evidentemente $P\left(\Omega\right) = 1$.

Mostremos que P es finitamente aditiva. En efecto, Si $E = \Pi_u^{-1}\left(A_u\right)$ y $F = \Pi_v^{-1}\left(B_v\right)$ son elementos de $\Im_0$, ajenos, sea $w = u \cup v$, $A_w = \Pi_{wu}^{-1}(A_u)$ y $B_w = \Pi_{wv}^{-1}(B_v)$, entonces A_w y B_w son ajenos y $E \cup F = \Pi_w^{-1}\left(A_w \cup B_w\right)$, así que:

$P\left(E \cup F\right) = P_w\left(A_w \cup B_w\right) = P_w\left(A_w\right) + P_w\left(B_w\right)$

$= P_u\left(A_u\right) + P_v\left(B_v\right) = P\left(E\right) + P\left(F\right)$.

Mostremos ahora que P es σ-subaditiva. Para esto, por el teorema 5.2, basta con demostrar que si tenemos una sucesión decreciente $\left(E_i\right)_{i\in\mathbb{N}}$, de elementos de $\Im_0$, tal que $\bigcap_{i=1}^{\infty} E_i = \emptyset$, entonces $\lim_{i\rightsquigarrow\infty} P\left(E_i\right) = 0$.

Para cada $i \in \mathbb{N}$, sea $v_i \in U$ y $A_i \in \mathcal{B}_{v_i}$ tales que $E_i = \Pi_{v_i}^{-1}\left(A_i\right)$, y definamos $u_i = \bigcup_{j=1}^{i} v_j$ y $B_i = \Pi_{u_i v_i}^{-1}(A_i)$. Entonces $B_i \in \mathcal{B}_{u_i}$, $E_i = \Pi_{u_i}^{-1}\left(B_i\right)$ y la sucesión de conjuntos $\left(u_n\right)_{n\in\mathbb{N}}$ es creciente.

Supongamos que $\varepsilon = \lim_{i\rightsquigarrow\infty} P\left(E_i\right) > 0$.

Por el teorema 15.4, para cada $i \in \mathbb{N}$, $P_{u_i}\left(B_i\right)$ puede ser aproximada por medidas de compactos contenidos en B_i, en particular existe un subconjunto compacto de $\mathbb{R}^{u_i}$, contenido en B_i, tal que:

$P_{u_i}\left(B_i\right) - P_{u_i}\left(K_i\right) < \frac{\varepsilon}{2^{i+1}}$.

Sea $F_i = \Pi_{u_i}^{-1}\left(K_i\right)$.

Obviamente se tiene $F_i \subset E_i$ para cualquier $i \in \mathbb{N}$, así que si demostramos que $\bigcap_{i=1}^{\infty} F_i \neq \emptyset$, habremos demostrado que $\bigcap_{i=1}^{\infty} E_i \neq \emptyset$, llegando así a una contradicción.

Para cada $j \in \mathbb{N}$ definamos $H_j = \bigcap_{i=1}^{j} F_i$. Entonces $H_j \subset \bigcap_{i=1}^{j} E_i = E_j$ y se tiene:

$P\left(E_j\right) - P\left(H_j\right) = P\left(E_j\right) - P\left(\bigcap_{i=1}^{j} F_i\right) = P\left(\bigcup_{i=1}^{j}\left(E_j - F_i\right)\right)$

$\leq \sum_{i=1}^{j} P\left(E_j - F_i\right) \leq \sum_{i=1}^{j} P\left(E_i - F_i\right) = \sum_{i=1}^{j}\left[P\left(E_i\right) - P\left(F_i\right)\right]$

$= \sum_{i=1}^{j} [P_{u_i}(B_i) - P_{u_i}(K_i)] < \sum_{i=1}^{j} \frac{\varepsilon}{2^{i+1}} < \frac{\varepsilon}{2}$.

Así que:

$P(H_j) > P(E_j) - \frac{\varepsilon}{2} \geq \frac{\varepsilon}{2} > 0$.

Por lo tanto $H_j \neq \emptyset$ para cualquier $j \in \mathbb{N}$.

Para cada $j \in \mathbb{N}$, sea $x^{(j)} \in H_j$, entonces $x^{(j)} \in F_i = \Pi_{u_i}^{-1}(K_i)$ para $i \in \{1, \ldots, j\}$. Por lo tanto, $x_{u_i}^{(j)} = \Pi_{u_i} x^{(j)} \in K_i$ para $i \in \{1, \ldots, j\}$.

Visto de otra manera, fijando $i \in \mathbb{N}$ se tiene $x_{u_i}^{(j)} \in K_i$ para cualquier $j \in \{i, i+1, \ldots\}$.

En particular, $\left(x_{u_1}^{(j)}\right)_{j\in\mathbb{N}}$ es una sucesión en K_1, así que tiene por lo menos una subsucesión convergente $\left(x_{u_1}^{(m_{1,j})}\right)_{j\in\mathbb{N}}$, donde se puede asumir que $m_{1,j} > 1$.

A su vez, $\left(x_{u_2}^{(m_{1,j})}\right)_{j\in\mathbb{N}}$ es una sucesión en K_2, así que tiene por lo menos una subsucesión convergente $\left(x_{u_2}^{(m_{2,j})}\right)_{j\in\mathbb{N}}$, donde se puede asumir que $m_{2,j} > 2$. Además, $\Pi_{u_2 u_1}(x_{u_2}^{(m_{2,j})}) = x_{u_1}^{(m_{1,j})}$, así que, si $x_{u_1} = \lim_{j\rightsquigarrow\infty} x_{u_1}^{(m_{1,j})}$ y $x_{u_2} = \lim_{j\rightsquigarrow\infty} x_{u_2}^{(m_{2,j})}$, entonces $\Pi_{u_2 u_1}(x_{u_2}) = x_{u_1}$.

Continuando de la misma forma, obtenemos, para cada $i \in \mathbb{N}$, una sucesión convergente $\left(x_{u_i}^{(m_{i,j})}\right)_{j\in\mathbb{N}}$ en K_i, de tal manera que, si $x_{u_i} = \lim_{j\rightsquigarrow\infty} x_{u_i}^{(m_{i,j})}$, entonces $\Pi_{u_{i+1} u_i}(x_{u_{i+1}}) = x_{u_i}$.

Hemos obtenido entonces una sucesión $(x_{u_i})_{i\in\mathbb{N}}$ tal que $x_{u_i} \in K_i$ para cualquier $i \in \mathbb{N}$ y si $i, j \in \mathbb{N}$, con $i < j$, entonces $\Pi_{u_j u_i}(x_{u_j}) = x_{u_i}$.

Hablando informalmente, si pegamos los elementos x_{u_i} de esta sucesión obtenemos un punto y en $\mathbb{R}^{\cup_{i=1}^{\infty} u_i}$; este punto es una función de $\cup_{i=1}^{\infty} u_i$ en $\mathbb{R}$ la cual está definida por:

$y(t) = x_{u_i}(t)$ si $t \in u_i$.

Denotando, para cada $j \in \mathbb{N}$, por $\Pi^{(u_j)}$ a la proyección de $\mathbb{R}^{\cup_{i=1}^{\infty} u_i}$ sobre $\mathbb{R}^{u_j}$, este elemento y así definido tiene la siguiente propiedad:

$\Pi^{(u_j)}(y) = x_{u_j} \in K_j$.

Para tener definido un punto en $\bigcap_{i=1}^{\infty} F_i$ únicamente resta completar y definiendo arbitrariamente las coordenadas que corresponden a $\Gamma - \cup_{i=1}^{\infty} u_i$; por ejemplo definamos $x \in \Omega$ de la siguiente manera:

$$x(t) = \begin{cases} y(t) & \text{si } t \in \cup_{i=1}^{\infty} u_i \\ 0 & \text{en otro caso} \end{cases}$$

Entonces $\Pi_{u_i}(x) = x_{u_i} \in K_i$ para cualquier $i \in \mathbb{N}$, así que $x \in \Pi_{u_i}^{-1}(K_i) = F_i$ para cualquier $i \in \mathbb{N}$, es decir:

$x \in \bigcap_{i=1}^{\infty} F_i \subset \bigcap_{i=1}^{\infty} E_i$.

lo cual establece la contradicción mencionada.

Por lo tanto $\lim_{i \rightsquigarrow \infty} P(E_i) = 0$, así que P es σ-subaditiva.

De acuerdo con el teorema de extensión de Carathéodory, existe una única medida de probabilidad P definida sobre $\Im = \sigma(\Im_0)$ tal que $P(\Pi_u^{-1}(B_u)) = P_u(B_u)$ para cualquier $u \in U$ y $B_u \in \mathcal{B}_u$.

Para $t \in \Gamma$, sea $X_t : \Omega \mapsto \mathbb{R}$ definida por $X_t(\omega) = \omega(t)$. Entonces $[X_t \leq x] = \Pi_{\{t\}}^{-1}((-\infty, x]) \in \Im_0$ para cualquier $t \in \Gamma$ y $x \in \mathbb{R}$, así que X_t es $\Im$-medible para cualquier $t \in \Gamma$. Además, para cualquier $u = \{t_1, \ldots, t_n\} \in U$ y $(x_1, \ldots, x_n) \in \mathbb{R}^u$, se tiene:

$P[X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n] = P(\Pi_u^{-1}((-\infty, x_1] \times \cdots \times (-\infty, x_n]))$

$= P_u((-\infty, x_1] \times \cdots \times (-\infty, x_n]) = F_u(x_1, \ldots, x_n)$.

Así que F_u es la función de distribución conjunta de $X_{t_1}, \ldots, X_{t_n}$.

$\blacksquare$

APÉNDICES

A.1. Teorema de Heine Borel

En el año 1895, Émile Borel encontró una propiedad de los intervalos cerrados y acotados en $\mathbb{R}$, la cual generó un concepto muy importante en el Análisis Matemático, el de conjunto compacto.

Borel estaba atacando un problema de continuación analítica de una función de variable compleja y, como parte de su razonamiento, demostró un resultado, el cual, simplificado para el caso de un intervalo $[a, b]$ de números reales, se puede enunciar como sigue:

Si $I_1, I_2, \ldots$ es una familia infinita numerable de intervalos abiertos tales que la suma de sus longitudes es menor que la longitud del intervalo $[a, b]$, entonces la unión de todos los intervalos I_n no cubre al intervalo $[a, b]$.

Para probar lo anterior, Borel demostró que si la unión de los intervalos I_n cubriera al intervalo $[a, b]$, entonces existiría un subcolección finita $I_{n_1}, I_{n_2}, \ldots, I_{n_m}$, de esos intervalos, cuya unión cubriría a $[a, b]$ y entonces se tendría:

$\sum_{n=1}^{\infty} l(I_n) \geq \sum_{k=1}^{m} l\left(I_{n_k}\right) \geq b - a.$

Años más tarde, se encontraría un resultado más general, al cual ahora se le conoce como teorema de Heine-Borel. Su demostración se encuentra un poco más adelante en esta sección.

Definición A.5. *Diremos que un subconjunto A de $\mathbb{R}^n$ está acotado si existe una bola abierta tal que A está contenido en ella.*

Definición A.6. *Llamaremos celda de $\mathbb{R}^n$ a un conjunto de la forma $I_1 \times I_2 \times \cdots \times I_n$, donde $I_1 = [a_1, b_1]$, $I_2 = [a_2, b_2], \cdots, I_n = [a_n, b_n]$ son intervalos cerrados y acotados de números reales tales que, para cualquier $j \in \{1, 2, \ldots, n\}$, $a_j < b_j$.*

Definición A.7. *Si $(C_m)_{m\in\mathbb{N}}$ es una sucesión de celdas de $\mathbb{R}^n$, diremos que éstas están anidadas si $C_{k+1} \subset C_k$ para cualquier $k \in \mathbb{N}$.*

Proposición A.4. *Sea $(C_m)_{m\in\mathbb{N}}$ una sucesión de celdas anidadas de $\mathbb{R}^n$, entonces $\cap_{m=1}^{\infty} C_m \neq \emptyset$.*

Demostración

Sea $I_1^{(m)} \times I_2^{(m)} \times \cdots \times I_n^{(m)}$ la celda C_m; entonces, para cada $j \in \{1, 2, \ldots, n\}$, los intervalos $I_j^{(1)}$, $I_j^{(2)}$, $I_j^{(3)}$, $\ldots$ forman una sucesión anidada de intervalos cerrados y acotados; por lo tanto, existe un número real x_j en la intersección de todos ellos. Evidentemente, para cualquier $m \in \mathbb{N}$, el vector $(x_1, x_2, \ldots, x_n)$ perternece a la celda C_m.

■

TEOREMA A.1. *Si K es un subconjunto de $\mathbb{R}^n$, cerrado y acotado, entonces, para cualquier familia infinita $\{G_\gamma : \gamma \in \Gamma\}$ de subconjuntos abiertos de $\mathbb{R}^n$ tales que $K \subset \bigcup_{\gamma \in \Gamma} G_\gamma$, existe un subconjunto U de Γ, finito, tal que $K \subset \bigcup_{u \in U} G_u$.*

Demostración

Sea $\{G_\gamma : \gamma \in \Gamma\}$ una familia de subconjuntos abiertos de $\mathbb{R}^n$ tales que $K \subset \bigcup_{\gamma \in \Gamma} G_\gamma$ y denotemos por $\mathbb{U}$ a la familia de todos los subconjuntos finitos Γ.

Supongamos que no existe algún conjunto $U \in \mathbb{U}$ tal que $K \subset \bigcup_{u \in U} G_u$.

Como K es acotado, existe una celda $I_1^{(1)} \times I_2^{(1)} \times \cdots \times I_n^{(1)}$ que lo contiene, a la cual llamaremos C_1. Podemos tomarla de tal forma que los intervalos $I_1, I_2, \ldots, I_n$ tengan la misma longitud, la cual denotaremos por L.

Vamos a construir, inductivamente, una sucesión $(C_m)_{m \in \{2,3,\ldots\}}$ de celdas tales que, para cualquier $m \in \{2, 3, \ldots\}$:

i) $C_m \subset C_{m-1}$.

ii) $K \cap C_m \neq \emptyset$ y no existe algún conjunto $U \in \mathbb{U}$ tal que $K \cap C_m \subset \bigcup_{u \in U} G_u$.

iii) Si $C_m = I_1^{(m)} \times I_2^{(m)} \times \cdots \times I_n^{(m)}$, entonces, para cualquier $j \in \{1, 2, \ldots, n\}$, $l\left(I_j^{(m)}\right) = \frac{1}{2^{(m-2)}} L$.

Definiendo $C_2 = C_1$, C_2 cumple con las condiciones i ii y iii.

Tomemos ahora $k \in \{2, 3, \ldots\}$ y supongamos que tenemos definida una celda $C_k = \left[a_1^{(k)}, b_1^{(k)}\right] \times \left[a_2^{(k)}, b_2^{(k)}\right] \times \cdots \times \left[a_n^{(k)}, b_n^{(k)}\right]$ satisfaciendo las propiedades i ii y iii.

Para cada $j \in \{1, 2, \ldots, n\}$, denotemos por $c_j^{(k)}$ al punto medio del intervalo $\left[a_j^{(k)}, b_j^{(k)}\right]$. De esta forma, en cada coordenada j tenemos los intervalos $\left[a_j^{(k)}, c_j^{(k)}\right]$ y $\left[c_j^{(k)}, b_j^{(k)}\right]$. Tomando en cada coordenada uno de esos dos intervalos y considerando el producto cartesiano de ellos, formamos una celda. El total de celdas que podemos formar de esa manera es igual a 2^n y si $C = I_1 \times I_2 \times \cdots \times I_n$ es cualquiera de esas celdas, se tiene $C \subset C_k$ y, para cualquier $j \in \{1, 2, \ldots, n\}$, se tiene:

$l(I_j) = \frac{1}{2} l\left(\left[a_j^{(k)}, b_j^{(k)}\right]\right) = \frac{1}{2}\frac{1}{2^{(k-2)}} L = \frac{1}{2^{(k-1)}} L.$

Sabemos que no existe algún conjunto $U \in \mathbb{U}$ tal que $K \cap C_k \subset \bigcup_{u\in U} G_u$, así que, por lo menos para una de las 2^n celdas que formamos, llamémosla C, se tiene que no existe algún conjunto $U \in \mathbb{U}$ tal que $K \cap C \subset \bigcup_{u\in U} G_u$, porque si para cualquiera de las 2^n celdas se tuviera la propiedad contraria, existiría un conjunto $U \in \mathbb{U}$ tal que $K \cap C_k \subset \bigcup_{u\in U} G_u$. Para esa celda C se tiene $K \cap C \neq \emptyset$ ya que de otra forma se tendría $K \cap C \subset G_\gamma$ para cualquier $\gamma \in \Gamma$, lo cual contradice la propiedad con la que elegimos a C.

Definamos entonces C_{k+1} como una cualquiera de esas celdas C, entre las 2^n celdas que formamos, con la propiedad de que no existe algún conjunto $U \in \mathbb{U}$ tal que $K \cap C \subset \bigcup_{u\in U} G_u$.

La celda C_{k+1} así definida satisface entonces las propiedades i, ii y iii.

Así que, por el principio de inducción matemática, para cada $m \in \{2, 3, \ldots\}$, queda definida cada una de las celdas C_m satisfaciendo las propiedades i, ii y iii.

Denotemos por $L^{(m)}$ a la longitud común, igual a $\frac{1}{2^{(m-2)}} L$, de cada uno de los intervalos $I_j^{(m)}$ que componen la celda C_m.

Por la propiedad i, las celdas de la sucesión $(C_m)_{m\in\mathbb{N}}$ que hemos construido están anidadas, así que $\cap_{m=1}^{\infty} C_m \neq \emptyset$. Esta intersección es un conjunto formado por un único punto, ya que si $x = (x_1, x_2, \ldots, x_n)$ y $y = (y_1, y_2, \ldots, y_n)$ pertenecen a esa intersección, entonces, para cada $j \in \{1, 2, \ldots, n\}$ y cualquier $m \in \{2, 3, \ldots\}$, x_j y y_j pertenecen al intervalo $I_j^{(m)}$ cuya longitud es igual a $L^{(m)}$; así que $|y_j - x_j| \leq L^{(m)} = \frac{1}{2^{(m-2)}} L$ para cualquier $m \in \{2, 3, \ldots\}$ y, entonces, $x_j = y_j$.

Sea $z = (z_1, z_2, \ldots, z_n)$ el único punto en la intersección $\cap_{m=1}^{\infty} C_m$.

Para cada $m \in \{2, 3, \ldots\}$, tomemos un elemento $z^{(m)} = \left(z_1^{(m)}, z_2^{(m)}, \ldots, z_n^{(m)}\right) \in K \cap C_m$, entonces tanto z como $z^{(m)}$ pertenecen a la celda C_m, así que, para cualquier $j \in \{1, 2, \ldots, n\}$, se tiene:

$\left| z_j^{(m)} - z_j \right| \leq L^{(m)}.$

Por lo tanto:

$d\left(z^{(m)}, z\right) = \sqrt{\left(z_1^{(m)} - z_1\right)^2 + \left(z_2^{(m)} - z_2\right)^2 + \cdots + \left(z_n^{(m)} - z_n\right)^2}$

$\leq L^{(m)}\sqrt{n} = \frac{1}{2^{(m-2)}} L\sqrt{n}.$

Así que la sucesión $\left(z^{(m)}\right)_{m\in\{2,3,\ldots\}}$ converge a z.

Además, $z^{(m)} \in K$ para cualquier $m \in \{2, 3, \ldots\}$, así que, como K es cerrado, $z = \lim_{m\to\infty} z^{(m)} \in K$.

Por hipótesis, $K \subset \bigcup_{\gamma\in\Gamma} G_\gamma$, así que, teniendo $z \in K$, existe algún conjunto G_{γ_0} tal que $z \in G_{\gamma_0}$. Siendo G_{γ_0} un conjunto abierto, existe una bola abierta, $B_r(z)$, con centro z y un radio positivo r tal que $B_r(z) \subset G_{\gamma_0}$.

Por otra parte, como z es el centro de la bola $B_r(z)$, tomando $h = \frac{r}{2\sqrt{n}}$, la celda $[z_1 - h, z_1 + h] \times [z_2 - h, z_2 + h] \times \cdots \times [z_n - h, z_n + h]$ está contenida en $B_r(z)$. En efecto, si:

$$x = (x_1, x_2, \ldots, x_n) \in [z_1 - h, z_1 + h] \times [z_2 - h, z_2 + h] \times \cdots \times [z_n - h, z_n + h],$$

entonces, para cada $j \in \{1, 2, \ldots, n\}$, se tiene $|x_j - z_j| \leq h$, así que:

$$d(x, z) = \sqrt{(x_1 - z_1)^2 + (x_2 - z_2)^2 + \cdots + (x_n - z_n)^2} \leq h\sqrt{n} = \tfrac{r}{2}.$$

Tomemos $m_0 \in \{2, 3, \ldots\}$ tal que $L^{(m_0)} = \frac{1}{2^{(m-2)}}L < h$, entonces, como:

$$z \in C_{m_0} = \left[a_1^{(m)}, b_1^{(m)}\right] \times \left[a_2^{(m)}, b_2^{(m)}\right] \times \cdots \times \left[a_n^{(m)}, b_n^{(m)}\right],$$

si $y = (y_1, y_2, \ldots, y_n)$ es cualquier elemento de C_{m_0} se tiene, para cualquier $j \in \{1, 2, \ldots, n\}$:

$$|y_j - z_j| \leq L^{(m_0)}.$$

Así que:

$$d(y, z) = \sqrt{(y_1 - z_1)^2 + (y_2 - z_2)^2 + \cdots + (y_n - z_n)^2} \leq L^{(m_0)}\sqrt{n} < h\sqrt{n} = \tfrac{r}{2}.$$

Por lo tanto, la celda C_{m_0} está contenida en la bola $B_r(z)$, la cual a su vez está contenida en G_{γ_0}. En particular, se tiene $K \cap C_{m_0} \subset G_{\gamma_0}$.

Hemos llegado a una contradicción ya que construimos la sucesión $(C_m)_{m\in\{2,3,\ldots\}}$ de tal forma que, para cualquier $m \in \{2, 3, \ldots\}$, no existe algún conjunto $U \in \mathbb{U}$ tal que $K \cap C_m \subset \bigcup_{u\in U} G_u$.

Por lo tanto, la hipótesis de la que partimos, a saber, que no existe algún conjunto $U \in \mathbb{U}$ tal que $K \subset \bigcup_{u\in U} G_u$, es falsa.

Así que existe algún conjunto $U \in \mathbb{U}$ tal que $K \subset \bigcup_{u\in U} G_u$, lo cual prueba el resultado.

■

Demostraremos ahora el inverso del teorema A.1.

TEOREMA A.2. *Sea K un subconjunto de $\mathbb{R}^n$ con la propiedad de que, para cualquier familia $\{G_\gamma : \gamma \in \Gamma\}$ de subconjuntos abiertos de $\mathbb{R}^n$ tales que $K \subset \bigcup_{\gamma\in\Gamma} G_\gamma$, existe un subconjunto U de Γ, finito, tal que $K \subset \bigcup_{u\in U} G_u$, entonces K es cerrado y acotado.*

Demostración

Tomemos un elemento cualquiera $z \in \mathbb{R}^n$. La familia de bolas abiertas $\{B_m(z) : m \in \mathbb{N}\}$ forman una cubierta de K, así que existe un subconjunto finito U de números naturales tales que $K \subset \bigcup_{m\in U} B_m(z)$. Si $m_0 = \max U$, entonces $\bigcup_{m\in U} B_m(z) = B_{m_0}(z)$, así que $K \subset B_{m_0}(z)$ y, por lo tanto, está acotado.

Para demostar que K es cerrado, tomemos un elemento cualquiera $y \in K^c$ y para cada $m \in \mathbb{N}$, denotemos por G_m al complemento de la bola cerrada $\overline{B}_{\frac{1}{m}}(y)$, de centro y y radio $\frac{1}{m}$. Se tiene entonces:

$$\bigcup_{m\in\mathbb{N}} G_m = \bigcup_{m\in\mathbb{N}} \overline{B}^c_{\frac{1}{m}}(y) = \left(\bigcap_{m\in\mathbb{N}} \overline{B}_{\frac{1}{m}}(y)\right)^c = (\{y\})^c = \mathbb{R}^n - \{y\}.$$

Así que, como $y \notin K$, entonces $K \subset \bigcup_{m\in\mathbb{N}} G_m$.

Como los conjuntos G_m son abiertos, existe un subconjunto finito U de números naturales tales que $K \subset \bigcup_{m\in U} G_m$. Si $m_0 = \max U$, entonces $\bigcup_{m\in U} G_m = G_{m_0}$, así que $K \subset G_{m_0}$. Por lo tanto, $B_{\frac{1}{m_0}}(y) \subset \overline{B}_{\frac{1}{m_0}}(y) = G^c_{m_0} \subset K^c$.

Así que, dado $y \in K^c$, existe una bola abierta de centro y, contenida en K^c; es decir, todos los puntos de K^c son interiores a K^c, así que K^c es cerrado y, por lo tanto, K es cerrado.

■

Combinando los teoremas A.1 y A.2, se tiene el siguiente resultado:

TEOREMA A.3 (Teorema de Heine Borel). *Un subconjunto K de $\mathbb{R}^n$ es cerrado y acotado si y sólo si, para cualquier familia infinita $\{G_\gamma : \gamma \in \Gamma\}$ de subconjuntos abiertos de $\mathbb{R}^n$ tales que $K \subset \bigcup_{\gamma\in\Gamma} G_\gamma$, existe un subconjunto U de Γ, finito, tal que $K \subset \bigcup_{u\in U} G_u$.*

La propiedad que tienen los conjuntos cerrados y acotados en $\mathbb{R}^n$, enunciada en el teorema de Heine Borel, es llamada compacidad. Es decir decimos que un subconjunto K de $\mathbb{R}^n$ es compacto si para cualquier familia infinita $\{G_\gamma : \gamma \in \Gamma\}$ de subconjuntos abiertos de X tales que $K \subset \bigcup_{\gamma\in\Gamma} G_\gamma$, existe un conjunto finito U de Γ, finito, tal que $K \subset \bigcup_{u\in U} G_u$.

En el caso de cualquier espacio métrico, no siempre los conjuntos cerrados y acotados son compactos. Por ejemplo consideremos el conjunto $\mathbf{F} = \{f : [c,d] \mapsto \mathbb{R} : f \text{ es acotada}\}$, donde c y d son dos números reales tales que $c < d$; si $f \in \mathbf{F}$, definamos $\|f\|_s = \sup\{|f(x)| : x \in [a,b]\}$ y si $f, g \in \mathbf{F}$, definamos $d_s(f,g) = \|g - f\|_s$. Sabemos que $(\mathbf{F}, d_s)$ es un espacio métrico (completo).

Sea $(x_n)_{n\in\mathbb{N}}$ una sucesión de puntos distintos en $[c,d]$ y, para cada $n \in \mathbb{N}$, definamos $f_n : [c,d] \to \mathbb{R}$ de la siguiente manera:

$$f_n(x) = \begin{cases} 1 & \text{si } x = x_n \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene $\|f_n\|_s = 1$ para cualquier $n \in \mathbb{N}$ y $d_s(f_n, f_m) = 1$ para cualquier pareja de números naturales, m y n, tales que $n \neq m$.

Evidentemente, el conjunto $K = \{f_1, f_2, \ldots\}$ es acotado.

Además, K no tiene puntos de acumulación. En efecto, ninguna de las funciones f_n puede ser punto de acumulación de K, ya que, dada cualquiera de ellas, la bola con centro en esa función y radio $\frac{1}{2}$ no contiene alguna otra función en K. Ahora, si una función f en $\mathbf{F}$, que no pertenece a K, fuera punto de acumulación de K, entonces existirían dos funciones f_{n_1} y f_{n_2} en K tales que:

$0 < d_s(f_{n_1}, f) < \frac{1}{4}$,

$0 < d_s(f_{n_2}, f) < d_s(f_{n_1}, f)$.

Por la última desigualdad, se tendría $f_{n_1} \neq f_{n_2}$, así que $d_s(f_{n_1}, f_{n_2}) = 1$.

Por otra parte, se tendría:

$d_s(f_{n_1}, f_{n_2}) \leq d_s(f_{n_1}, f) + d_s(f, f_{n_2}) < 2d_s(f_{n_1}, f) < \frac{1}{2}$,

llegando así a una contradicción.

Siendo vacío el conjunto de puntos de acumulación de K, podemos concluir que K es cerrado.

Tenemos entonces que el conjunto K es cerrado y acotado.

Ahora bien, consideremos, para cada $n \in \mathbb{N}$, la bola abierta de radio $\frac{1}{4}$ y centro f_n. La unión de esas bolas contiene a K, pero la unión de cualquier colección finita de esas bolas únicamente contiene un número finito de elementos de K, a saber, los centros de ellas.

Por lo tanto, K es un conjunto cerrado y acotado que no es compacto.

A.2. Conjuntos compactos

En esta sección (X, d) será un espacio métrico.

DEFINICIÓN A.8. *Diremos que $K \subset X$ es compacto si para cualquier familia infinita $\{G_\gamma : \gamma \in \Gamma\}$ de subconjuntos abiertos de X tales que $K \subset \bigcup_{\gamma \in \Gamma} G_\gamma$, existe un conjunto finito $T \subset \Gamma$ tal que $K \subset \bigcup_{\gamma \in T} G_\gamma$.*

DEFINICIÓN A.9. *Diremos que $K \subset X$ es numerablemente compacto si para cualquier familia $\{G_n : n \in \mathbb{N}\}$ de subconjuntos abiertos de X tales que $K \subset \bigcup_{n \in \mathbb{N}} G_n$, existe un conjunto finito $T \subset \mathbb{N}$ tal que $K \subset \bigcup_{n \in T} G_n$.*

DEFINICIÓN A.10. *Diremos que $K \subset X$ es secuencialmente compacto si para toda sucesión $(x_n)_{n \in \mathbb{N}}$ de elementos de K existe una subsucesión que converge a algún elemento de K.*

DEFINICIÓN A.11. *Diremos que una familia de subconjuntos de X, $\{F_\gamma : \gamma \in \Gamma\}$, tiene la propiedad de la intersección finita si dado cualquier subconjunto finito $T \subset \Gamma$, se tiene $\bigcap_{\gamma \in T} F_\gamma \neq \emptyset$.*

DEFINICIÓN A.12. *Diremos que $A \subset X$ es acotado si existe una bola abierta que lo contiene.*

DEFINICIÓN A.13. *Diremos que $A \subset X$ es totalmente acotado si para cualquier $\varepsilon > 0$, existe un conjunto finito de bolas cerradas de radio ε cuya unión cubre A.*

Obviamente, todo conjunto totalmente acotado es acotado, sin embargo, el inverso no siempre es verdadero. Por ejemplo, consideremos nuevamente el conjunto:

$\mathbf{F} = \{f : [c, d] \mapsto \mathbb{R} : f \text{ es acotada}\}$,

donde c y d son dos números reales tales que $c < d$; y tomemos en $\mathbf{F}$ la norma de la convergencia uniforme $\|\|_s$.

Consideremos el subconjunto de $\mathbf{F}$ formado por las funciones f_n $(n \in \mathbb{N})$ tales que el conjunto $K = \{f_1, f_2, \ldots\}$ es cerrado y acotado pero no compacto.

El conjunto K no es totalmente acotado; en efecto, como $\|f_n\|_s = 1$ para cualquier $n \in \mathbb{N}$ y $d_s(f_n, f_m) = 1$ para cualquier pareja de números naturales, m y n, tales que $n \neq m$, entonces para $\varepsilon = \frac{1}{4}$, cualquier bola cerrada de radio ε contiene a lo más un elemento de K, ya que si f y g pertenecen a esa bola y h es su centro, entonces:

$d(f, g) \leq d(f, h) + d(h, g) \leq \frac{1}{2}$.

Por lo tanto, no existe un conjunto finito de bolas cerradas de radio $\varepsilon = \frac{1}{4}$ cuya unión cubra K.

En $\mathbb{R}^n$, si un subconjunto A es acotado, entonces es totalmente acotado. En efecto, siendo A acotado existe una bola abierta B que lo contiene; por lo tanto, la bola cerrada $\overline{B}$ también lo contiene. Ahora bien, siendo la bola cerrada un conjunto compacto, es totalmente acotado; de manera que entonces A también es totalmente acotado.

La demostración de la proposición A.2 no utiliza propiedades particulares de $\mathbb{R}^n$ y se aplica tanto a los conjuntos compactos como a los numerablemente compactos; así que se tienen los siguientes resultados:

PROPOSICIÓN A.5. *Si $K \subset X$ es un conjunto compacto, entonces es cerrado y acotado.*

PROPOSICIÓN A.6. *Si $K \subset X$ es un conjunto numerablemente compacto, entonces es cerrado y acotado.*

PROPOSICIÓN A.7. *Si $K \subset X$ es un conjunto secuencialmente compacto, entonces es cerrado y acotado.*

Demostración

Si K no fuera acotado, dada cualquier bola abierta $B_r(x)$, de centro $x \in X$ y radio $r \in (0, \infty)$, existiría algún elemento de K que no pertenecería a esa bola.

Suponiendo entonces que K no es acotado, definamos, inductivamente, una sucesión $(x_i)_{i \in \mathbb{N}}$ de elementos de K y una sucesión $(r_i)_{i \in \mathbb{N}}$ de números reales positivos tales que :

i) $x_i \in K - B_{r_i}\left(\overline{0}\right)$.

i) $r_i = d\left(x_{i-1}, \overline{0}\right) + 1$.

donde tomamos $x_0 = \overline{0}$.

Definamos $r_1 = 1$ y x_1 como cualquiera de los elementos de K que no pertenecen a la bola $B_{r_1}\left(\overline{0}\right)$. Obviamente, x_1 y r_1 satisfacen las propiedades i y ii.

Tomemos ahora $k \in \mathbb{N}$ y supongamos que tenemos definidos x_k y r_k satisfaciendo las propiedades i y ii.

Definamos $r_{k+1} = d\left(x_k, \overline{0}\right) + 1$ y x_{k+1} como cualquiera de los elementos de K que no pertenecen a la bola $B_{r_{k+1}}\left(\overline{0}\right)$. Obviamente, x_{k+1} y r_{k+1} satisfacen las propiedades i y ii.

Así que, por el principio de inducción matemática, quedan definidas las sucesiones $(x_i)_{i \in \mathbb{N}}$ y $(r_i)_{i \in \mathbb{N}}$ satisfaciendo las propiedades i y ii.

Para cualquier $j \in \mathbb{N}$, se tiene:

$d\left(x_{j+1}, \overline{0}\right) \geq r_{j+1} = d\left(x_j, \overline{0}\right) + 1.$

Así que, si x_j y x_{j+k} son dos elementos cualesquiera de la sucesión $(x_i)_{i \in \mathbb{N}}$, donde j y k son números naturales, se tiene:

$d\left(x_{j+k}, \overline{0}\right) \geq d\left(x_{j+k-1}, \overline{0}\right) + 1 \geq d\left(x_{j+k-2}, \overline{0}\right) + 2 \geq \cdots \geq d\left(x_j, \overline{0}\right) + k.$

Además:

$d\left(x_{j+k}, \overline{0}\right) \leq d\left(x_{j+k}, x_j\right) + d\left(x_j, \overline{0}\right).$

Así que:

$d\left(x_{j+k}, x_j\right) \geq d\left(x_{j+k}, \overline{0}\right) - d\left(x_j, \overline{0}\right) \geq d\left(x_j, \overline{0}\right) + k - d\left(x_j, \overline{0}\right) = k.$

Por lo tanto, la distancia entre dos elementos cualesquiera de la sucesión $(x_i)_{i \in \mathbb{N}}$ es mayor o igual a 1, así que no existe alguna subsucesión convergente de esa sucesión.

Por otra parte, si K no tiene puntos de acumulación, entonces es cerrado.

Si el conjunto de puntos de acumulación de K es no vacío, sea x cualquiera de esos puntos, entonces existe una sucesión $(x_i)_{i\in\mathbb{N}}$ de elementos de K que converge a x; por lo tanto cualquier subsucesión de $(x_i)_{i\in\mathbb{N}}$ converge también a x. Pero, por la hipótesis de la proposición, existe una subsucesión de $(x_i)_{i\in\mathbb{N}}$ que converge a algún elemento que pertenece a K; por lo tanto $x \in K$. Así que, también en este caso, K es un conjunto cerrado.

■

Proposición A.8. *Sea $\{K_n\}_{n\in\mathbb{N}}$ una familia de subconjuntos numerablemente compactos con la propiedad de la intersección finita, entonces $\bigcap_{n=1}^{\infty} K_n \neq \emptyset$.*

Demostración

Supongamos $\bigcap_{n=1}^{\infty} K_n = \emptyset$, entonces $\bigcup_{n=1}^{\infty} K_n^c = X$. Así que $K_{n_0} \subset \bigcup_{n=1}^{\infty} K_n^c$ para cualquier $n_0 \in \mathbb{N}$. Sea $n_0 \in \mathbb{N}$ arbitraria, entonces, como K_{n_0} es compacto, existen $n_1 \ldots, n_m \in \mathbb{N}$ tales que $K_{n_0} \subset \bigcup_{k=1}^{m} K_{n_k}^c$. Entonces $\bigcap_{k=0}^{m} K_{n_k} = K_{n_0} \cap \left(\bigcup_{k=1}^{m} K_{n_k}^c\right)^c = \emptyset$, lo cual es una contradicción.

■

Proposición A.9. *Sea $\{K_\gamma\}_{\gamma\in\Gamma}$ una familia de subconjuntos compactos con la propiedad de la intersección finita, entonces $\bigcap_{\gamma\in\Gamma} K_\gamma \neq \emptyset$.*

Demostración

Supongamos $\bigcap_{\gamma\in\Gamma} K_\gamma = \emptyset$, entonces $\bigcup_{\gamma\in\Gamma} K_\gamma^c = X$. Así que $K_{\gamma_0} \subset \bigcup_{\gamma\in\Gamma} K_\gamma^c$ para cualquier $\gamma_0 \in \Gamma$. Sea $\gamma_0 \in \Gamma$ arbitraria, entonces, como K_{γ_0} es compacto, existen $\gamma_1 \ldots, \gamma_n \in \Gamma$ tales que $K_{\gamma_0} \subset \bigcup_{k=1}^{n} K_{\gamma_k}^c$. Entonces $\bigcap_{k=0}^{n} K_{\gamma_k} = K_{\gamma_0} \cap \left(\bigcup_{k=1}^{n} K_{\gamma_k}^c\right)^c = \emptyset$, lo cual es una contradicción.

■

Proposición A.10. *Sea $K \subset X$ cerrado. Supongamos que cualquier familia de subconjuntos cerrados de K con la propiedad de la intersección finita tiene una intersección no vacía. Entonces K es compacto.*

Demostración

Sea $\{G_\gamma\}_{\gamma\in\Gamma}$ una familia de subconjuntos abiertos tales que $K \subset \bigcup_{\gamma\in\Gamma} G_\gamma$ y S la familia de los subconjuntos finitos de Γ.

Para cada $T \in S$, definamos $E_T = \bigcup_{\gamma\in T} G_\gamma$.

Si $K \cap E_T^c \neq \emptyset$ para cualquier $T \in S$, entonces la familia de conjuntos cerrados $\{K \cap E_T^c\}_{T\in S,}$ tiene la propiedad de la intersección finita. Por lo tanto, $K \cap \left(\bigcup_{\gamma\in\Gamma} G_\gamma\right)^c = \bigcap_{\gamma\in\Gamma} (K \cap E_T^c) \neq \emptyset$, lo cual es una contradicción.

Por lo tanto, $K \cap E_T^c = \emptyset$ para algún $T \in S$, así que $K \subset \bigcup_{\gamma\in T} G_\gamma$ para algún $T \in S$.

Así que K es compacto. ■

Proposición A.11. *Sea $K \subset X$ cerrado. Supongamos que cualquier familia numerable de subconjuntos cerrados de K con la propiedad de la intersección finita tiene una intersección no vacía. Entonces K es numerablemente compacto.*

Demostración

Sea $\{G_n\}_{n\in\mathbb{N}}$ una familia de subconjuntos abiertos tales que $K \subset \bigcup_{n=1}^{\infty} G_n$ y, para cada $n \in \mathbb{N}$, definamos $E_n = \bigcup_{k=1}^{n} G_k$.

Si $K \cap E_n^c \neq \emptyset$ para cualquier $n \in \mathbb{N}$, entonces la familia de conjuntos cerrados $\{K \cap E_n^c\}_{n\in\mathbb{N}}$ tiene la propiedad de la intersección finita. Por lo tanto, $K \cap (\bigcup_{n=1}^{\infty} G_n)^c = \bigcap_{n=1}^{\infty} (K \cap E_n^c) \neq \emptyset$, lo cual es una contradicción.

Por lo tanto, $K \cap E_n^c = \emptyset$ para alguna $n \in \mathbb{N}$, así que $K \subset \bigcup_{k=1}^{n} G_k$ para alguna $n \in \mathbb{N}$.

Así que K es numerablemente compacto.

■

Corolario A.3. *Un conjunto cerrado $K \subset X$ es compacto si y sólo si cualquier familia de subconjuntos cerrados de K con la propiedad de la intersección finita tiene una intersección no vacía.*

Corolario A.4. *Un conjunto cerrado $K \subset X$ es numerablemente compacto si y sólo si cualquier familia numerable de subconjuntos cerrados de K con la propiedad de la intersección finita tiene una intersección no vacía.*

Teorema A.4. *Un conjunto $K \subset X$ es secuencialmente compacto si y sólo si es numerablemente compacto.*

Demostración

Supongamos que K es secuencialmente compacto y sea $\{F_n\}_{n\in\mathbb{N}}$ familia de subconjuntos cerrados de K con la propiedad de la intersección finita.

Para cada $n \in \mathbb{N}$, definamos $H_n = \bigcap_{k=1}^{n} F_k$ y tomemos $x_n \in H_n$.

Obsérvese que, para cualquier $j \in \mathbb{N}$, $x_n \in F_j$ para cualquier $n \geq j$.

Como K es secuencialmente compacto, existe una subsucesión convergente, $(x_{n_k})_{k\in\mathbb{N}}$, de $(x_n)_{n\in\mathbb{N}}$.

Sea $x = \lim_{k\to\infty} x_{n_k}$.

Si $j \in \mathbb{N}$ y $k \geq j$, entonces $n_k \geq j$, así que $x_{n_k} \in F_j$. Por lo tanto, $x \in \bigcap_{k=1}^{\infty} F_k$.

Así que K es numerablemente compacto.

Inversamente, supongamos que K es numerablemente compacto y sea $(x_n)_{n\in\mathbb{N}}$ una sucesión de elementos de K.

Para cada $n \in \mathbb{N}$, definamos $C_n = \{x_n, x_{n+1}, \ldots\}$. Entonces, la familia de conjuntos cerrados $\left\{\overline{C_n}\right\}_{n\in\mathbb{N}}$ tiene la propiedad de la intersección finita. Por lo tanto, $\bigcap_{n=1}^{\infty} \overline{C_n} \neq \emptyset$.

Sea $x \in \bigcap_{n=1}^{\infty} \overline{C_n}$, entonces, para cada $n \in \mathbb{N}$, existe $x_m \in C_n$ tal que $d(x_m, x) < \frac{1}{n}$. Podemos definir entonces inductivamente una subsucesión $(x_{n_k})_{k\in\mathbb{N}}$ de $(x_n)_{n\in\mathbb{N}}$ tal que $d(x_{n_k}, x) < \frac{1}{n_k}$ para cualquier $k \in \mathbb{N}$. Así que $(x_{n_k})_{k\in\mathbb{N}}$ converge a x.

Por lo tanto, K es secuencialmente compacto.

■

Proposición A.12. *Sea $K \subset X$ un conjunto secuencialmente compacto. Entonces, para cualquier $\varepsilon > 0$, existe un conjunto finito $T \subset K$ tal que $K \subset \bigcup_{x\in T} B_\varepsilon(x)$.*

Demostración

Supongamos que para alguna $\varepsilon > 0$ se tiene que $K \cap \left(\bigcup_{x\in T} B_\varepsilon(x)\right)^c \neq \emptyset$ para cualquier conjunto finito $T \subset K$. Sea $x_1 \in K$ arbitrario y definamos inductivamente una sucesión $(x_n)_{n\in\mathbb{N}}$ de elementos de K tal que $x_{k+1} \in \bigcap_{j=1}^{k} B_\varepsilon^c(x_j)$, es decir, $d(x_{k+1}, x_j) \geq \varepsilon$ para cualquier $j \in \{1, \ldots, k\}$. Se tiene entonces $d(x_n, x_m) \geq \varepsilon$ para cualesquiera $n, m \in \mathbb{N}$ distintas. Así que no existe ninguna subsucesión convergente de $(x_n)_{n\in\mathbb{N}}$, lo cual es una contradicción ya que K es secuencialmente compacto.

■

Corolario A.5. *Si $K \subset X$ es un conjunto secuencialmente compacto, entonces es totalmente acotado.*

Corolario A.6. *Sea $K \subset X$ un conjunto secuencialmente compacto. Entonces K contiene un subconjunto denso numerable.*

Demostración

Para cada $n \in \mathbb{N}$, sea T_n un subconjunto finito de K tal que $K \subset \bigcup_{x\in T_n} B_{\frac{1}{n}}(x)$. Definamos $T = \bigcup_{n=1}^{\infty} T_n$.

Si $x \in K$ entonces, para cada $n \in \mathbb{N}$, existe $x_n \in T_n \subset T \subset K$ tal que $d(x_n, x) < \frac{1}{n}$. Así que $x \in \overline{T}$, es decir, $K = \overline{T}$.

■

Proposición A.13. *Si $A \subset X$ es totalmente acotado y $H = \{G_\gamma : \gamma \in \Gamma\}$ es una familia de subconjuntos abiertos de X cuya unión cubre A, entonces existe una colección numerable de conjuntos $G\gamma \in H$ cuya unión sigue cubriendo A.*

Demostración

Si A es vacío el resultado es trivial; así que asumiremos que $A \neq \emptyset$.

Para cada $n \in \mathbb{N}$, sea B_n un subconjunto finito de X tal que la unión de las bolas cerradas de centro cada uno de los puntos de B_n y radio $\frac{1}{n}$ cubre A y sea $B = \cup_{n=1}^{\infty} B_n$.

Definamos:

$M = \Big\{(n, y) : n \in \mathbb{N},\ y \in B_n \text{ y existe } \gamma \in \Gamma \text{ tal que } \overline{B}_{\frac{1}{n}}(y) \subset G_\gamma\Big\}.$

Como $A \subset \bigcup_{\gamma \in T} G_\gamma$, dado $x \in A$, existe $\gamma \in \Gamma$ tal que $x \in G_\gamma$. Siendo G_γ abierto, existe una bola $B_r(x)$ contenida en G_γ. Tomemos $n \in \mathbb{N}$ tal que $\frac{1}{n} < \frac{r}{2}$.

Como $A \subset \bigcup_{y \in B_n} \overline{B}_{\frac{1}{n}}(y)$, existe $y \in B_n$ tal que $x \in \bar{B}_{\frac{1}{n}}(y)$. Entonces, si $z \in \overline{B}_{\frac{1}{n}}(y)$, se tiene:

$d(z, x) \leq d(z, y) + d(y, x) \leq \frac{1}{n} + \frac{1}{n} = \frac{2}{n} < r.$

Así que $z \in B_r(x)$; por lo tanto:

$\overline{B}_{\frac{1}{n}}(y) \subset B_r(x) \subset G_\gamma.$

Hemos demostrado entonces que, dado $x \in A$, existen $n \in \mathbb{N}$ y $y \in B_n$ tales que $(n, y) \in M$ y $x \in \bar{B}_{\frac{1}{n}}(y)$.

En particular, lo anterior demuestra que el conjunto M es no vacío y, obviamente, es un conjunto numerable.

Denotemos por (r_1, s_1), (r_2, s_2), $\ldots$ los elementos de M.

Para cada $(r_k, s_k) \in M$, tomemos un elemento $\gamma \in \Gamma$ tal que $\bar{B}_{\frac{1}{r_k}}(s_k) \subset G_\gamma$ y denotemos ese elemento por γ_k.

Con esta notación, podemos enunciar el resultado anterior de la siguiente manera:

Para cada $x \in A$, existe un elemento (r_k, s_k) tal que $x \in \bar{B}_{\frac{1}{r_k}}(s_k) \subset G_{\gamma_k}$.

Por lo tanto, $A \subset \bigcup_k G_{\gamma_k}$, lo cual demuestra el resultado

■

Corolario A.7. *Un conjunto $K \subset X$ es secuencialmente compacto si y sólo si es compacto.*

Demostración

Si K es compacto, entonces es numerablemente compacto, así que, por el teorema A.4, es secuencialmente compacto.

Inversamente, si K es secuencialmente compacto, entonces, por el teorema A.4, es numerablemente compacto. Además, por el corolario A.5, también es totalmente acotado, así que, por la proposición A.13, si $H = \{G_\gamma : \gamma \in \Gamma\}$ es una familia de subconjuntos abiertos de X cuya unión cubre A, entonces existe un conjunto numerable $\Psi \subset \Gamma$ tal que $K \subset \bigcup_{\gamma \in \Psi} G_\gamma$. Por lo tanto, siendo K numerablemente compacto, existe un conjunto finito $T \subset \Psi$ tal que $K \subset \bigcup_{\gamma \in T} G_n$. Lo cual demuestra que K es compacto.

También como corolario, se tiene el siguiente resultado:

TEOREMA A.5. *Si K es un subconjunto de X, las siguientes condiciones son equivalentes:*

(i) *K es compacto.*
(ii) *K es numerablemente compacto.*
(iii) *K es secuencialmente compacto.*

A.3. Caracterización de los conjuntos compactos

DEFINICIÓN A.14. *Diremos que $A \subset X$ es relativamente compacto si $\overline{A}$ es compacto.*

Por la proposición A.5, el corolario A.7 y el corolario A.5, se tiene el siguiente resultado:

PROPOSICIÓN A.14. *Si $K \subset X$ es un conjunto compacto, entonces es cerrado y totalmente acotado.*

PROPOSICIÓN A.15. *Un conjunto $B \subset X$ es totalmente acotado si y sólo si para toda sucesión $(x_n)_{n\in\mathbb{N}}$ de elementos de B contiene una subsucesión de Cauchy.*

Demostración

Supongamos que B es totalmente acotado y sea $(y_n)_{n\in\mathbb{N}}$ una sucesión de elementos de B.

Tomemos un conjunto finito $T_1 \subset X$ tal que $K \subset \bigcup_{x\in T_1} B_1(x)$. Siendo T_1 finito, por lo menos una de las bolas $B_1(x)$, con $x \in T_1$, contiene una infinidad de elementos de la sucesión $(y_n)_{n\in\mathbb{N}}$. Sea $B_1(x_1)$ una de esas bolas.

Tomemos ahora un conjunto finito $T_2 \subset X$ tal que $K \subset \bigcup_{x\in T_2} B_{\frac{1}{2}}(x)$. Siendo T_2 finito, por lo menos una de las bolas $B_{\frac{1}{2}}(x)$, con $x \in T_2$, es tal que $B_1(x_1) \cap B_{\frac{1}{2}}(x)$ contiene una infinidad de elementos de la sucesión $(y_n)_{n\in\mathbb{N}}$. Sea $B_{\frac{1}{2}}(x_2)$ una de esas bolas.

Mediante ese proceso podemos definir inductivamente una sucesión de bolas $\left\{B_{\frac{1}{n}}(x_n)\right\}_{n\in\mathbb{N}}$ tales que, para cualquier $n \in \mathbb{N}$, el conjunto $\bigcap_{k=1}^{n} B_{\frac{1}{k}}(x_k)$ contiene una infinidad de elementos de la sucesión $(y_n)_{n\in\mathbb{N}}$.

Tomemos $y_{n_1} \in B_1(x_1)$ y definamos inductivamente una subsucesión $(y_{n_k})_{k\in\mathbb{N}}$ tal que, para cualquier $k \in \mathbb{N}$, $y_{n_k} \in \bigcap_{j=1}^{k} B_{\frac{1}{j}}(x_j)$. Se tiene entonces $d(y_{n_k}, x_j) < \frac{1}{j}$ para cualquier $j \in \{1, \ldots, k\}$; así que, fijando $j \in \mathbb{N}$, se tiene $d(y_{n_k}, x_j) < \frac{1}{j}$ para cualquier $k \geq j$.

Por lo tanto, fijando $j \in \mathbb{N}$, se tiene $d(y_{n_k}, y_{n_k}) < \frac{2}{j}$ para cualesquiera $k, m \geq j$. Así que la sucesión $(y_{n_k})_{k\in\mathbb{N}}$ es de Cauchy.

Supongamos ahora que B no es totalmente acotado. Entonces, existe $\varepsilon > 0$ tal que $K \cap \left(\bigcup_{x\in T} B_\varepsilon(x)\right)^c \neq \emptyset$ para cualquier conjunto finito $T \subset X$ (en particular, para cualquier conjunto finito $T \subset K$). Sea $x_1 \in K$ arbitrario y definamos inductivamente una sucesión $(x_n)_{n\in\mathbb{N}}$ de elementos de K tal que $x_{k+1} \in \bigcap_{j=1}^{k} B_\varepsilon^c(x_j)$, es decir, $d(x_{k+1}, x_j) \geq \varepsilon$ para cualquier $j \in \{1, \ldots, k\}$. Se tiene entonces $d(x_n, x_m) \geq \varepsilon$ para cualesquiera $n, m \in \mathbb{N}$ distintas. Así que no existe ninguna subsucesión de $(x_n)_{n\in\mathbb{N}}$ que sea de Cauchy.

■

Proposición A.16. *Un conjunto $B \subset X$ es relativamente compacto si y sólo si para toda sucesión $(x_n)_{n\in\mathbb{N}}$ de elementos de B existe una subsucesión convergente.*

Demostración

Si B es relativamente compacto, entonces $\overline{B}$ es secuencialmente compacto, así que toda sucesión $(y_n)_{n\in\mathbb{N}}$ de elementos de B contiene una subsucesión convergente (a algún punto de $\overline{B}$).

Inversamente, supongamos que para toda sucesión $(x_n)_{n\in\mathbb{N}}$ de elementos de B existe una subsucesión convergente y sea $(y_n)_{n\in\mathbb{N}}$ una sucesión de elementos de $\overline{B}$.

Para cada $n \in \mathbb{N}$, sea $z_n \in B$ tal que $d(y_n, z_n) < \frac{1}{n}$. Tal z_n existe pues si $y_n \in B$ podemos tomar $z_n = y_n$ y si $y_n \notin B$ entonces y_n es punto de acumulación de B.

Sea ahora $(z_{n_k})_{k\in\mathbb{N}}$ una subsucesión convergente de $(z_n)_{n\in\mathbb{N}}$ y $z = \lim_{k\rightsquigarrow\infty} z_{n_k}$. Entonces, como:

$d(y_{n_k}, z) \leq d(y_{n_k}, z_{n_k}) + d(z_{n_k}, z) < d(z_{n_k}, z) + \frac{1}{n_k}$,

se tiene $z = \lim_{k\rightsquigarrow\infty} y_{n_k}$.

Además, como $\overline{B}$ es cerrado, $z \in \overline{B}$.

Por lo tanto, para toda sucesión $(y_n)_{n\in\mathbb{N}}$ de elementos de $\overline{B}$ existe una subsucesión convergente a algún elemento de $\overline{B}$. Es decir, $\overline{B}$ es secuencialmente compacto y, por lo tanto, compacto.

■

Corolario A.8. *Si X es completo, entonces un conjunto $K \subset X$ es relativamente compacto si y sólo si es totalmente acotado.*

Demostración

Supongamos que K es relativamente compacto y tomemos una sucesión $(x_n)_{n\in\mathbb{N}}$ cualquiera de elementos de K; entonces, por la proposición A.16, existe una subsucesión de $(x_n)_{n\in\mathbb{N}}$ que es convergente y, por lo tanto, de Cauchy; así que por la proposición A.15, K es totalmente acotado.

Inversamente, supongamos que K es totalmente acotado y tomemos una sucesión $(x_n)_{n\in\mathbb{N}}$ cualquiera de elementos de K ; entonces, por la proposición A.15, existe una subsucesión de

$(x_n)_{n\in\mathbb{N}}$ que es de Cauchy. Siendo X completo, esa subsucesión es convergente; así que, por la proposición A.16, K es relativamente compacto.

■

Corolario A.9. *Si X es completo, entonces cualquier conjunto $K \subset X$ cerrado y totalmente acotado, es compacto.*

Demostración

Si $K \subset X$ es un conjunto cerrado y totalmente acotado, entonces, por el corolario A.8, es relativamente compacto; así que $K = \overline{K}$ es compacto.

■

Por la proposición A.14 y el corolario A.9 se tiene entonces el siguiente resultado:

Teorema A.6. *Si X es completo, un conjunto $K \subset X$ es compacto si y sólo si es cerrado y totalmente acotado.*

Si X no es completo, es posible que haya subconjuntos cerrrados y totalmente acotados que no sean compactos. Por ejemplo, tomemos $X = \mathbb{Q}$ con la distancia usual entre números reales; entonces el conjunto $A = \{x \in \mathbb{Q} : \ 2 < x^2 < 3\}$ es cerrado y totalmente acotado, pero no compacto.

A.4. Espacios vectoriales normados

Asumimos que el lector está familiarizado con las propiedades básicas de los espacios vectoriales sobre un campo.

Definición A.15. *Sea X un espacio vectorial sobre $\mathbb{R}$. Diremos que una función $x \to \|x\|$, definida sobre X y con valores en $\mathbb{R}$, es una norma si satisface las siguientes propiedades:*

(i) $\|x\| \geq 0$ *para cualquier* $x \in X$.
(ii) $\|x\| = 0$ *si y sólo si* $x = 0$.
(iii) $\|\alpha x\| = |\alpha| \, \|x\|$ *para cualesquiera* $\alpha = F$ *y* $x \in X$.
(iv) $\|x + y\| \leq \|x\| + \|y\|$ *para cualesquiera* $x, y \in X$.

Recordemos que a partir de una norma se puede definir una métrica de la siguiente manera:

$d(x, y) = \|y - x\|$.

Definición A.16. *Si X es un espacio vectorial en donde está definida una norma, diremos que X es un espacio vectorial normado.*

Definición A.17. *Sea X un espacio vectorial de dimensión finita sobre $\mathbb{R}$ y $B = \{x_1, \ldots, x_n\}$ una base de X. Si $x = \sum_{k=1}^{n} \alpha_k x_k$, definimos $\|x\|_{0,B} = \text{máx}\{|\alpha_k| : k \in \{1, \ldots, n\}\}$.*

Se prueba inmediatamente que la función $x \mapsto \|x\|_{0,B}$ es una norma sobre X.

DEFINICIÓN A.18. *La norma $\|x\|_{0,B}$ será llamada la norma del máximo en la base B.*

Además si $\|\|$ es cualquier otra norma definida sobre X y $x = \sum_{k=1}^{n} \alpha_k x_k$, entonces:

$\|x\| = \|\sum_{k=1}^{n} \alpha_k x_k\| \leq \sum_{k=1}^{n} |\alpha_k| \|x_k\| \leq \|x\|_{0,B} \sum_{k=1}^{n} \|x_k\|$.

PROPOSICIÓN A.17. *Sea X un espacio vectorial de dimensión finita sobre $\mathbb{R}$ y $B = \{x_1, \ldots, x_n\}$ una base de X. Entonces X es completo con respecto a la norma $\|x\|_{0,B}$.*

Demostración

Sea $(y_m)_{m\in\mathbb{N}}$ una sucesión de Cauchy con respecto a $\|\|_{0,B}$. Si $y_m = \sum_{k=1}^{n} \alpha_k^{(m)} x_k$, entonces, como $\left|\alpha_k^{(i)} - \alpha_k^{(j)}\right| \leq \|y_i - y_j\|_{0,B}$ para cualesquiera $i, j \in \mathbb{N}$, la sucesión $\left(\alpha_k^{(m)}\right)_{m\in\mathbb{N}}$ es de Cauchy para cualquier $k \in \{1, \ldots, n\}$. Sea $\alpha_k = \text{lím}_{m\rightsquigarrow\infty} \alpha_k^{(m)}$, para $k \in \{1, \ldots, n\}$, y definamos $y = \sum_{k=1}^{n} \alpha_k x_k$.

Dada $\varepsilon > 0$, sea N tal que $\|y_i - y_j\|_{0,B} < \varepsilon$ para cualesquiera $i, j \in \mathbb{N}$ mayores o iguales a N. Entonces, también se tiene $\left|\alpha_k^{(i)} - \alpha_k^{(j)}\right| < \varepsilon$ para cualesquiera $i, j \in \mathbb{N}$ mayores o iguales a N y cualquier $k \in \{1, \ldots, n\}$. Fijando $i \geq N$, se tiene, para cualquier $k \in \{1, \ldots, n\}$:

$\left|\alpha_k^{(i)} - \alpha_k\right| = \text{lím}_{j\rightsquigarrow\infty} \left|\alpha_k^{(i)} - \alpha_k^{(j)}\right| \leq \varepsilon.$

Por lo tanto:

$\|y_i - y\|_{0,B} = \text{máx}\left\{\left|\alpha_k^{(i)} - \alpha_k\right| : k \in \{1, \ldots, n\}\right\} \leq \varepsilon.$

Así que la sucesión $(y_m)_{m\in\mathbb{N}}$ converge a y.

■

PROPOSICIÓN A.18. *Sea X un espacio vectorial de dimensión finita sobre $\mathbb{R}$, $B = \{x_1, \ldots, x_n\}$ una base de X y $\|\|$ una norma definida sobre X. Entonces existe una constante positiva c tal que $\|x\| \geq c\|x\|_{0,B}$ para cualquier $x \in X$.*

Demostración

La demostración se hará por inducción sobre la dimensión de X.

Supongamos que el resultado es válido para cualquier espacio vectorial de dimensión k, cualquier base de ese espacio y cualquier norma definida sobre él.

Sea ahora Y un espacio vectorial de dimensión $k+1$, $B_Y = \{y_1, \ldots, y_{k+1}\}$ una base de Y y $\|\|$ una norma definida sobre Y.

Tomemos cualquier $y_i \in B_Y$ y sea M el subespacio vectorial generado por $A = B_Y - \{y_i\}$.

Por la hipótesis de inducción, sabemos que existe una constante positiva c_M tal que $\|y\| \geq c_M \|y\|_{0,A}$ para cualquier $y \in M$.

Sea $(z_m)_{m\in\mathbb{N}}$ una sucesión de Cauchy, en M, con respecto a la norma $\|\|$. Entonces, como $\|y\|_{0,A} \leq \frac{1}{c_M}\|y\|$ para cualquier $y \in M$, $(z_m)_{m\in\mathbb{N}}$ es también una sucesión de Cauchy con respecto a la norma $\|\|_{0,A}$. Sea z el límite de la sucesión con respecto a $\|\|_{0,A}$. Entonces, como $\|z_m - z\| \leq \|z_m - z\|_{0,A}\left(\sum_{j=1}^{k+1}\|y_j\| - \|y_i\|\right)$ para cualquier $m \in \mathbb{N}$, $(z_m)_{m\in\mathbb{N}}$ también converge a z con respecto a la norma $\|\|$. Así que M es completo y, por lo tanto, es un subconjunto cerrado de X.

Como los vectores $y_1, \ldots, y_{k+1}$ son linealmente independientes, $y_i \in M^c$. Así que existe una bola abierta de radio $\delta_i > 0$ y centro y_i, completamente contenida en M^c. Por tanto, $\|y - y_i\| \geq \delta_i$ para cualquier $y \in M$.

Tomemos $y = \sum_{j=1}^{k+1}\alpha_j y_j \in Y$.

Si $\alpha_i \neq 0$, entonces $y_i - \sum_{j=1}^{k+1}\frac{\alpha_j}{\alpha_i}y_j \in M$, así que $\left\|\sum_{j=1}^{k+1}\frac{\alpha_j}{\alpha_i}y_j\right\| \geq \delta_i$. Por lo tanto:

$$\|y\| = \left\|\sum_{j=1}^{k+1}\alpha_j y_j\right\| \geq |\alpha_i|\,\delta_i.$$

Obviamente también se tiene $\|y\| \geq |\alpha_i|\,\delta_i$ cuando $\alpha_i = 0$.

Definamos $c_Y = \text{mín}\{\delta_i : i \in \{1, \ldots, k+1\}\}$. Entonces $c_Y > 0$ y se tiene $\|y\| \geq |\alpha_i|\,c_Y$ para cualquier $i \in \{1, \ldots, k+1\}$. Así que:

$$\|y\| \geq c_Y \,\text{máx}\{|\alpha_i| : \{\delta_i : i \in \{1, \ldots, k+1\}\}\} = c_Y\|y\|_{0,B_Y}.$$

■

Corolario A.10. *Sea X un espacio vectorial de dimensión finita sobre $\mathbb{F}$ y $\|\|_1$ y $\|\|_2$ dos normas definidas sobre X. Entonces existen dos constantes positivas a y b tales que*

$$a\|x\|_1 \leq \|x\|_2 \leq b\|x\|_1$$

para cualquier $x \in X$.

Corolario A.11. *Sea X un espacio vectorial de dimensión finita sobre $\mathbb{F}$. Entonces X es completo con respecto a cualquier norma definida sobre él.*

Corolario A.12. *Sea X un espacio vectorial normado. Entonces cualquier subespacio de X de dimensión finita es cerrado.*

Demostración

Todo subespacio vectorial de dimensión finita es completo. Por lo tanto, es un subconjunto cerrado de X. ■

Proposición A.19. *Sea X un espacio vectorial normado de dimensión finita. Entonces todo subconjunto de X, cerrado y acotado, es compacto.*

Demostración

Sea $B = \{x_1, \ldots, x_m\}$ una base de X y a, b dos constantes positivas tales que $\|x\|_{0,B} \leq a \|x\|$ y $\|x\| \leq b \|x\|_{0,B}$ para cualquier $x \in X$, en donde $\|\|$ es la norma en X.

Sea $K \subset X$ un conjunto cerrado y acotado, M tal que $\|x\| \leq M$ para cualquier $x \in K$ y $(y_n)_{n\in\mathbb{N}}$ una sucesión de elementos en K.

Si, para cada $n \in \mathbb{N}$, $y_n = \sum_{i=1}^{m} \alpha_{ni} x_i$, entonces:

$\max\{|\alpha_{ni}| : i \in \{1, \ldots, n\}\} = \|y_n\|_{0,B} \leq a \|y_n\| \leq aM.$

Así que, para cada $i \in \{1, \ldots, n\}$, la sucesión $(\alpha_{ni})_{n\in\mathbb{N}}$ está acotada.

Sean $\left(\alpha_{n_k^{(1)}1}\right)_{k\in\mathbb{N}}$ una subsucesión convergente de $(\alpha_{n1})_{n\in\mathbb{N}}$, $\left(\alpha_{n_k^{(2)}2}\right)_{k\in\mathbb{N}}$ una subsucesión convergente de $\left(\alpha_{n_k^{(1)}2}\right)_{k\in\mathbb{N}}$, ... Obtenemos de esta forma una subsucesión $(y_{n_k})_{k\in\mathbb{N}}$ de $(y_n)_{n\in\mathbb{N}}$ tal que, para cualquier $i \in \{1, \ldots, n\}$, la sucesión $(\alpha_{n_k i})_{k\in\mathbb{N}}$ es convergente.

Para $i \in \{1, \ldots, n\}$, sea $\alpha_i = \lim_{k\rightsquigarrow\infty} \alpha_{n_k i}$ y definamos $y = \sum_{i=1}^{m} \alpha_i x_i$. Entonces:

$\lim_{k\rightsquigarrow\infty} \|y_{n_k} - y\| \leq \lim_{k\rightsquigarrow\infty} b \|y_{n_k} - y\|_{0,B}$

$= b \lim_{k\rightsquigarrow\infty} \max\{|\alpha_{n_k i} - \alpha_i| : i \in \{1, \ldots, n\}\} = 0.$

Por lo tanto, la subsucesión $(y_{n_k})_{k\in\mathbb{N}}$ es convergente.

Además, como K es cerrado, $y \in K$.

K es entonces secuencialmente compacto y, por lo tanto, compacto.

■

Lema A.2. *Sea X un espacio vectorial normado, M un subespacio vectorial cerrado, contenido propiamente en X, y $\varepsilon \in (0, 1)$ arbitraria. Entonces existe un vector $x_\varepsilon \in X$ de norma 1 y tal que $\|x_\varepsilon - x\| > 1 - \varepsilon$ para cualquier $x \in M$.*

Demostración

Sea $y \in X - M$ y definamos:

$d = \inf\{\|y - x\| : x \in M\}.$

d es positiva pues si d fuera igual a cero entonces existiría una sucesión de elementos de M que converge a y, así que, como M es cerrado, se tendría $y \in M$.

Ahora bien, como $d < \frac{d}{1-\varepsilon}$, existe $z \in M$ tal que $d \leq \|y - z\| < \frac{d}{1-\varepsilon}$.

Definamos $x_\varepsilon = \frac{y-z}{\|y-z\|}$. Entonces $\|x_\varepsilon\| = 1$ y, si $x \in M$, se tiene:

$$\|x_\varepsilon - x\| = \left\| \frac{y-z}{\|y-z\|} - x \right\| = \frac{1}{\|y-z\|} \|(y-z) - \|y-z\| \, x\|$$

$$= \frac{1}{\|y-z\|} \|y - (z + \|y-z\| \, x)\| \geq \frac{d}{\|y-z\|} > 1 - \varepsilon.$$

■

Proposición A.20. *Sea X un espacio vectorial normado de dimensión infinita. Entonces la bola cerrada de radio 1 y centro 0 no es un conjunto compacto.*

Demostración

Sea $x_1 \in X$ un vector arbitrario de norma 1. Utilizando el lema A.2, podemos definir inductivamente una sucesión de vectores $x_n \in X$ tales que, para cualquier $n \in \mathbb{N}$, $\|x_n\| = 1$ y $\|y - x_{n+1}\| > \frac{1}{2}$ para cualquier y en el espacio vectorial generado por $\{x_1, \ldots, x_n\}$, el cual es cerrado por ser de dimensión finita. En particular, se tiene $\|x_n - x_m\| > \frac{1}{2}$ para cualesquiera $n, m \in \mathbb{N}$ distintos.

Supongamos que existe una subsucesión $(x_{n_k})_{k \in \mathbb{N}}$ convergente. Tal sucesión sería de Cauchy, así que existiría N tal que $\left\| x_{n_j} - x_{n_k} \right\| < \frac{1}{2}$ para cualesquiera $j, k \in \mathbb{N}$ mayores que N, lo cual es imposible.

Por lo tanto, la bola cerrada de radio 1 y centro 0 no es un conjunto secuencialmente compacto. Así que no es compacto.

■

Corolario A.13. *Sea X un espacio vectorial normado. Entonces X tiene dimensión finita si y sólo si todo subconjunto de X, cerrado y acotado, es compacto.*

Definición A.19. *Sea X un espacio vectorial normado. Diremos que X es un espacio de Banach si es completo con respecto a la métrica definida por la norma.*

Proposición A.21. *No existe ningún espacio de Banach con una base infinita numerable.*

Demostración

Sea X un espacio de Banach y supongamos que $B = \{x_1, x_2, \ldots\}$ es una base infinita numerable de X.

Para cada $n \in \mathbb{N}$, definamos $B_n = \{x_1, \ldots, x_n\}$ y denotemos por X_n al espacio vectorial generado por B_n.

Cada espacio vectorial X_n tiene dimensión finita, así que es completo. Por lo tanto, X_n es un subconjunto cerrado de X.

Sea $x \in X_n$, $\delta > 0$, $y \in B - B_n$ y $N \in \mathbb{N}$ tal que $\frac{1}{N} \|y\| < \delta$. Entonces, si $z = x + \frac{1}{N} y$, se tiene $z \notin X_n$ y $\|z - y\| = \frac{1}{N} \|y\| < \delta$.

Así que cualquier bola con centro x contiene elementos de X_n^c. Por lo tanto, el interior de X_n es vacío.

X_n es entonces un conjunto denso en ninguna parte.

Además, como B es una base de X, cualquier $x \in X$ pertenece a X_n para alguna $n \in \mathbb{N}$. Así que $X = \bigcup_{n=1}^{\infty} X_n$.

X es entonces de categoría I, lo cual nos conduce a una contradicción pues todo espacio métrico completo es de categoría II.

■

PROPOSICIÓN A.22. *Sea X un espacio vectorial normado y $(x_n)_{n\in\mathbb{N}}$ una sucesión de elementos de X tal que la serie $\sum_{n=1}^{\infty} \|x_n\|$ converge. Entonces la sucesión $(s_n)_{n\in\mathbb{N}}$ definida por $s_n = \sum_{k=1}^{n} x_k$ es de Cauchy.*

Demostración

Dada $\varepsilon > 0$, sea N tal que $\sum_{k=n}^{\infty} \|x_k\| < \varepsilon$ para cualquier $n \geq N$. Entonces, si $m > n \geq N$, se tiene:

$\|s_m - s_n\| = \left\|\sum_{k=n+1}^{m} x_k\right\| \leq \sum_{k=n+1}^{m} \|x_k\| \leq \sum_{k=n+1}^{\infty} \|x_k\| < \varepsilon.$

■

PROPOSICIÓN A.23. *Sea X un espacio vectorial normado. Supongamos que, para cualquier sucesión $(x_n)_{n\in\mathbb{N}}$ de elementos de X tal que la serie $\sum_{n=1}^{\infty} \|x_n\|$ converge, la sucesión $(s_n)_{n\in\mathbb{N}}$ definida por $s_n = \sum_{k=1}^{n} x_k$ converge. Entonces X es completo.*

Demostración

Sea $(x_n)_{n\in\mathbb{N}}$ una sucesión de Cauchy en X.

Definamos $m_0 = 0$ y, para cada $k \in \mathbb{N}$, tomemos $m_k \in \mathbb{N}$ tal que $m_k > m_{k-1}$ y $||x_n - x_m|| < \frac{1}{2^k}$ para cualesquiera $n, m \geq m_k$. Entonces, en particular se tiene $\left|\left|x_{m_{k+1}} - x_{m_k}\right|\right| < \frac{1}{2^k}$ para cualquier $k \in \mathbb{N}$. Así que la serie $\sum_{k=1}^{\infty} \left|\left|x_{m_{k+1}} - x_{m_k}\right|\right|$ converge. Por lo tanto, la sucesión $(s_n)_{n\in\mathbb{N}}$ definida por $s_n = x_{m_{n+1}} - x_{m_1} = \sum_{k=1}^{n} \left(x_{m_{k+1}} - x_{m_k}\right)$ converge. Así que la sucesión $(x_{m_n})_{n\in\mathbb{N}}$ también converge. Finalmente, tratándose de una sucesión de Cauchy, es suficiente la convergencia de una subsucesión para que la sucesión original converja.

■

A.5. Convergencia uniforme

DEFINICIÓN A.20. *Si $\mathbb{E}$ es cualquier conjunto, $D \subset \mathbb{E}$ y $(f_n)_{n\in\mathbb{N}}$ es una sucesión de funciones $f_n : \mathbb{E} \to \mathbb{R}$, diremos que $(f_n)_{n\in\mathbb{N}}$ converge uniformemente a la función $f : \mathbb{E} \to \mathbb{R}$, sobre el conjunto D, si dada cualquier $\varepsilon > 0$, existe $N \in \mathbb{N}$ tal que:*

$|f_n(x) - f(x)| < \varepsilon$

para cualquier $n \geq N$ *y* $x \in D$.

Un caso que de especial importancia es cuando D es un intervalo Sean a y b dos números reales tales que $a < b$ y $\mathbb{G} = \{f : [c, d] \mapsto \mathbb{R} : f \text{ es acotada}\}$. Si $f \in \mathbb{G}$, definimos:

$\|f\|_s = \sup\{|f(x)| : x \in [a, b]\}$.

PROPOSICIÓN A.24. *La función que asocia a cada* $f \in \mathbb{G}$ *el número real* $\|f\|_s$*, es una norma sobre* $\mathbb{G}$.

Demostración

Obviamente $\|f\|_s$ es un número real no negativo para cualquier $f \in \mathbb{G}$.

Si $\|f\|_s = 0$, entonces $\sup\{|f(x)| : x \in [a, b]\} = 0$, así que $|f(x)| = 0$ para cualquier $x \in [a, b]$; es decir, f es idénticamente 0.

Si f y g son elementos de $\mathbb{G}$, entonces:

$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_s + \|g\|_s$, para cualquier $x \in [a, b]$; así que:

$\|f + g\|_s = \sup\{|f(x) + g(x)| : x \in [a, b]\} \leq \|f\|_s + \|g\|_s$.

■

Si $f, g \in \mathbb{G}$ y definimos $d_s(f, g) = \|g - f\|_s$, entonces d_s es una métrica sobre $\mathbb{G}$.

PROPOSICIÓN A.25. $(\mathbb{G}, d_s)$ *es un espacio métrico completo.*

Demostración

Sea $(f_n)_{n \in \mathbb{N}}$ una sucesión de Cauchy en $\mathbb{G}$.

Dada $\varepsilon > 0$, tomemos $N \in \mathbb{N}$ tal que $\|f_n - f_m\|_s < \varepsilon$ para cualesquiera $n, m \in \mathbb{N}$ tales que $n \geq N$ y $m \geq N$.

Como $\|f_n - f_m\|_s = \sup\{|f_n(x) - f_m(x)| : x \in [a, b]\}$, se tiene:

$|f_n(x) - f_m(x)| < \varepsilon$ para cualquier $x \in [a, b]$ y cualesquiera $n, m \in \mathbb{N}$ tales que $n \geq N$ y $m \geq N$.

Así que, para cualquier $x \in [a, b]$, $(f_n(x))_{n \in \mathbb{N}}$ es una una sucesión de Cauchy de números reales; por lo tanto converge.

Definamos $f : [a, b] \mapsto \mathbb{R}$ de la siguiente manera:

$f(x) = \lim_{n \to \infty} f_n(x)$.

Como la sucesión $(f_n)_{n\in\mathbb{N}}$ es de Cauchy, está acotada, así que existe un número real M tal que $\|f_n\|_s \leq M$ para cualquier $n \in \mathbb{N}$. Por lo tanto:

$|f_n(x)| \leq M$ para cualquier $x \in [a,b]$ y cualquier $n \in \mathbb{N}$.

Así que, $|f(x)| \leq M$ para cualquier $x \in [a,b]$; es decir, f es acotada.

Dada $\varepsilon > 0$, tomemos $N \in \mathbb{N}$ tal que $\|f_n - f_m\|_s < \varepsilon$ para cualesquiera $n, m \in \mathbb{N}$ tales que $n \geq N$ y $m \geq N$. Entonces:

$|f_n(x) - f_m(x)| < \varepsilon$ para cualquier $x \in [a,b]$ y cualesquiera $n, m \in \mathbb{N}$ tales que $n \geq N$ y $m \geq N$. O bien:

$f_m(x) - \varepsilon < f_n(x) < f_m(x) + \varepsilon$ para cualquier $x \in [a,b]$ y cualesquiera $n, m \in \mathbb{N}$ tales que $n \geq N$ y $m \geq N$.

Por lo tanto, tomando límites cuando m tiende a infinito, se obtiene:

$f(x) - \varepsilon \leq f_n(x) \leq f(x) + \varepsilon$ para cualquier $x \in [a,b]$ y cualquier $n \geq N$. Es decir:

$|f_n(x) - f(x)| \leq \varepsilon$ para cualquier $x \in [a,b]$ y cualquier $n \geq N$.

Por lo tanto:

$\|f_n - f\|_s \leq \varepsilon$ para cualquier $n \geq N$.

Así que, la sucesión $(f_n)_{n\in\mathbb{N}}$ converge a f en el espacio métrico $(\mathbf{F}, d_s)$.

∎

Definición A.21. *Diremos que una función $f \in \mathbb{G}$ es lineal por pedazos si existe una partición $\{x_0, x_1, \ldots, x_n\}$ del intervalo $[a,b]$ tal que, para cualquier $j \in \{1, 2, \ldots, n\}$, la función f es lineal en el intervalo (x_{j-1}, x_j).*

Proposición A.26. *El conjunto $\mathbf{L}$, de funciones continuas y lineales por pedazos en $\mathbf{F}$, es denso en el conjunto $\mathbf{C}$, de funciones continuas en $\mathbf{F}$.*

Demostración

Si $f : [a,b] \mapsto \mathbb{R}$ es una función continua, entonces es uniformemente continua, así que, dada $\varepsilon > 0$, existe $\delta > 0$ tal que, si $x, y \in [a,b]$ y $|y - x| < \delta$, entonces $|f(y) - f(x)| < \frac{1}{2}\varepsilon$.

Consideremos una partición $\{x_0, x_1, \ldots, x_n\}$ una partición del intervalo $[a,b]$ tal que $x_i - x_{i-1} < \delta$ para cualquier $i \in \{1, 2, \ldots, n\}$ y definamos la función $f_\varepsilon : [a,b] \mapsto \mathbb{R}$ de la siguiente manera:

$$f_\varepsilon(x) = \begin{cases} f(x_{i-1}) + \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}(x - x_{i-1}) & \text{si } x \in (x_{i-1}, x_i) \text{ para alguna } i \in \{1, 2, \ldots, n\} \\ f(x) & \text{si } x \in \{x_0, x_1, \ldots, x_n\} \end{cases}$$

Si $x \in (x_{i-1}, x_i)$, se tiene:

$|f_\varepsilon(x) - f(x)| \leq |f_\varepsilon(x) - f(x_{i-1})| + |f(x_{i-1}) - f(x)|$

$= \left| \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} (x - x_{i-1}) \right| + |f(x_{i-1}) - f(x)| = |f(x_i) - f(x_{i-1})| \frac{x - x_{i-1}}{x_i - x_{i-1}} + |f(x_{i-1}) - f(x)|$

$\leq |f(x_i) - f(x_{i-1})| + |f(x_{i-1}) - f(x)| < \epsilon.$

Además, $|f_\varepsilon(x) - f(x)| = 0$ para cualquier $x \in \{x_0, x_1, \ldots, x_n\}$. Así que:

$|f_\varepsilon(x) - f(x)| < \epsilon$ para cualquier $x \in [a, b]$.

Por lo tanto:

$\|f_\varepsilon - f\|_s \leq \varepsilon.$

Se puede concluir entonces que, en particular, la sucesión de funciones $\left(f_{\frac{1}{n}}\right)_{n \in \mathbb{N}}$ converge a f en el espacio métrico $(\mathbf{F}, d_s)$.

■

A.6. Los racionales diádicos

Recordemos que los racionales diádicos son los números racionales de la forma $\frac{j}{2^n}$, donde $n \in \{0, 1, \ldots\}$ y j es un número entero. Recordemos también que el conjunto de racionales diádicos en un intervalo (a, b), es denso en $[a, b]$.

Definición A.22. *Denotaremos por $\mathbb{B}$ al conjunto:*

$\left\{(s_k)_{k \in \mathbb{N}} : s_k \in \{0, 1\} \text{ y } s_k = 1 \textit{ para una infinidad de índices } k\right\},$

por $\mathbb{B}_n$ al conjunto $\{(s_1, s_2, \ldots, s_n) : s_j \in \{0, 1\}$ para cualquier $j \in \{1, 2, \ldots, n\}\}$, donde $n \in \mathbb{N}$, y, para cada elemento $(s_1, s_2, \ldots, s_n) \in \mathbb{B}_n$, denotaremos por $I_{(s_1, s_2, \ldots, s_n)}$ al intervalo $\left(\sum_{k=1}^n \frac{s_k}{2^k}, \sum_{k=1}^n \frac{s_k}{2^k} + \frac{1}{2^n}\right]$.

Obsérvese que $I_{(s_1, s_2, \ldots, s_n)} \subset (0, 1]$ para cualquier $(s_1, s_2, \ldots, s_n) \in \mathbb{B}_n$. En efecto, para cualquier $(s_1, s_2, \ldots, s_n) \in \mathbb{B}_n$, se tiene:

$0 \leq \sum_{k=1}^n \frac{s_k}{2^k} \leq \sum_{k=1}^n \frac{1}{2^k} = 1 - \frac{1}{2^n}$

Proposición A.27. *1. Si $(s_1, s_2, \ldots s_n)$ y $(r_1, r_2, \ldots r_n)$ son dos elementos de $\mathbb{B}_n$, distintos, entonces los intervalos $I_{(s_1, s_2, \ldots, s_n)}$ y $I_{(r_1, r_2, \ldots, r_n)}$ son ajenos.*

2. $\bigcup_{\{(r_1, r_2, \ldots, r_n) \in \mathbb{B}_n\}} I_{(r_1, r_2, \ldots, r_n)} = (0, 1]$.

3. La función $f : \mathbb{B} \to (0, 1]$ definida por:

$f\left((s_k)_{k \in \mathbb{N}}\right) = \sum_{k=1}^\infty \frac{s_k}{2^k}$

es biyectiva.

4. Para cualquier $n \in \mathbb{N}$, si $(r_1, r_2, \ldots, r_n) \in \mathbb{B}_n$, entonces:

$$f^{-1}\left(I_{(s_1,s_2,\ldots,s_n)}\right) = \left\{(s_k)_{k\in\mathbb{N}} \in \mathbb{B}: \ s_k = r_k \text{ para cualquier } k \in \{1, 2, \ldots, n\}\right\}$$

Demostración

Cada punto $x \in (0,1]$, tiene un desarrollo único en base 2, el cual se obtiene de la siguiente manera:

Para iniciar el desarrollo, expresamos el intervalo $(0,1]$ como la unión de los intervalos ajenos $\left(0, \frac{1}{2}\right]$ y $\left(\frac{1}{2}, 1\right]$. Si $x \in \left(0, \frac{1}{2}\right]$, tomamos 0 como primer elemento de su desarrollo, mientras que si $x \in \left(\frac{1}{2}, 1\right]$, tomamos 1 como primer elemento de su desarrollo.

Para cada $n \in \mathbb{N}$, expresemos el intervalo $(0,1]$ como la unión de los intervalos ajenos $\left(0, \frac{1}{2^n}\right], \left(\frac{1}{2^n}, \frac{2}{2^n}\right], \cdots, \left(\frac{2^n-1}{2^n}, \frac{2^n}{2^n}\right]$. El punto x pertenece a uno y sólo uno de esos intervalos. Sea $\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]$ el intervalo al cual pertenece x. Si $k+1$ es un número impar, tomamos 0 como n-simo elemento de su desarrollo; si no lo es, tomamos 1 como n-simo elemento de su desarrollo.

Por la forma en que se define el desarrollo de x, si $s_1, s_2, \ldots$ son los términos de su desarrollo, se tiene:

$$x = \sum_{i=1}^{\infty} \frac{s_i}{2^i}$$

Hecho el desarrollo de esta manera, ningún punto $x \in (0,1]$ tiene un desarrollo $0.s_1 s_2 \cdots$ para el cual exista $N \in \mathbb{N}$ tal que $s_k = 0$ para cualquier $k \in \{N+1, N+2, \ldots\}$. En efecto, si lo hubiera, entonces x pertenecería a un intervalo de la forma $\left(\frac{k}{2^{N+1}}, \frac{k+1}{2^{N+1}}\right]$, donde $k+1$ es un número impar, digamos $k+1 = 2j-1$, con $j \in \mathbb{N}$; así que tendríamos $x \in \left(\frac{2j-2}{2^{N+1}}, \frac{2j-1}{2^{N+1}}\right]$; por lo tanto, o bien $x \in \left(\frac{4j-4}{2^{N+2}}, \frac{4j-3}{2^{N+2}}\right]$ o $x \in \left(\frac{4j-3}{2^{N+2}}, \frac{4j-2}{2^{N+2}}\right]$; pero, como término $N+2$ del desarrollo de x es 0, entonces $x \in \left(\frac{4j-4}{2^{N+2}}, \frac{4j-3}{2^{N+2}}\right]$. Repitiendo el razonamiento, llegaríamos a que $x \in \left(\frac{8j-8}{2^{N+3}}, \frac{8j-7}{2^{N+3}}\right]$. Y continuando con este procedimiento llegaríamos a que:

$$x \in \left(\frac{2^i(j-1)}{2^{N+i}}, \frac{2^i(j-1)+1}{2^{N+i}}\right] = \left(\frac{j-1}{2^N}, \frac{j-1}{2^{N+i}} + \frac{1}{2^{N+i}}\right]$$

para cualquier $i \in \mathbb{N}$.

Así que:

$$x \in \bigcap_{i=1}^{\infty} \left(\frac{j-1}{2^N}, \frac{j-1}{2^N} + \frac{1}{2^{N+i}}\right] = \emptyset$$

lo cual es una contradicción.

Como corolario se tiene que el desarrollo de cualquier número $x \in (0,1]$ contiene una infinidad de 1's.

Además, por lo anterior, f es una función suprayectiva de $\mathbb{B}$ en el intervalo $(0,1]$.

Por otra parte, para cualquier $n \in \mathbb{N}$, si $s_1, s_{2,} \ldots s_n$ son los primeros n términos del desarrollo de $x \in (0,1]$, entonces, como el desarrollo de x tiene una infinidad de 1's, se tiene:

$x > \sum_{i=1}^{n} \frac{s_i}{2^i}$

Además:

$x \leq \sum_{i=1}^{n} \frac{s_i}{2^i} + \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \sum_{i=1}^{n} \frac{s_i}{2^i} + \frac{1}{2^n}$

Por lo tanto:

$x \in \left(\sum_{i=1}^{n} \frac{s_i}{2^i}, \sum_{i=1}^{n} \frac{s_i}{2^i} + \frac{1}{2^n}\right]$

Sea ahora $(r_1, r_{2,} \ldots r_n)$ un elemento de $\mathbb{B}_n$, distinto de $(s_1, s_{2,} \ldots s_n)$, y sea $y \in (0,1]$ tal que los primeros n términos de su desarrollo son $r_1, r_{2,} \ldots r_n$.

Definamos:

$k_0 = \text{mín}\,\{k \in \{1, 2, \ldots, n\} : r_k \neq s_k\}$

Entonces:

$x \in \left(\sum_{i=1}^{k_0} \frac{s_i}{2^i}, \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{1}{2^{k_0}}\right]$

$y \in \left(\sum_{i=1}^{k_0} \frac{r_i}{2^i}, \sum_{i=1}^{k_0} \frac{r_i}{2^i} + \frac{1}{2^{k_0}}\right]$

$\sum_{i=1}^{k_0} \frac{r_i}{2^i} - \sum_{i=1}^{k_0} \frac{s_i}{2^i} = \frac{r_{k_0}}{2^{k_0}} - \frac{s_{k_0}}{2^{k_0}} = \frac{r_{k_0} - s_{k_0}}{2^{k_0}}$

Así que:

$\sum_{i=1}^{k_0} \frac{r_i}{2^i} = \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{r_{k_0} - s_{k_0}}{2^{k_0}}$

Por lo tanto:

$\left(\sum_{i=1}^{k_0} \frac{r_i}{2^i}, \sum_{i=1}^{k_0} \frac{r_i}{2^i} + \frac{1}{2^{k_0}}\right] = \left(\sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{r_{k_0} - s_{k_0}}{2^{k_0}}, \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{r_{k_0} - s_{k_0}}{2^{k_0}} + \frac{1}{2^{k_0}}\right]$

Así que:

$$\left(\sum_{i=1}^{k_0} \frac{r_i}{2^i}, \sum_{i=1}^{k_0} \frac{r_i}{2^i} + \frac{1}{2^{k_0}}\right] = \begin{cases} \left(\sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{1}{2^{k_0}}, \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{2}{2^{k_0}}\right] & \text{si } r_{k_0} - s_{k_0} = 1 \\ \left(\sum_{i=1}^{k_0} \frac{s_i}{2^i} - \frac{1}{2^{k_0}}, \sum_{i=1}^{k_0} \frac{s_i}{2^i}\right] & \text{si } r_{k_0} - s_{k_0} = -1 \end{cases}$$

En cualquiera de los dos casos $\left(\sum_{i=1}^{k_0} \frac{r_i}{2^i}, \sum_{i=1}^{k_0} \frac{r_i}{2^i} + \frac{1}{2^{k_0}}\right]$ y $\left(\sum_{i=1}^{k_0} \frac{s_i}{2^i}, \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{1}{2^{k_0}}\right]$ son ajenos.

Además:

$\left(\sum_{i=1}^{n} \frac{s_i}{2^i}, \sum_{i=1}^{n} \frac{s_i}{2^i} + \frac{1}{2^n}\right] \subset \left(\sum_{i=1}^{k_0} \frac{s_i}{2^i}, \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{1}{2^{k_0}}\right]$

$\left(\sum_{i=1}^{n} \frac{r_i}{2^i}, \sum_{i=1}^{n} \frac{r_i}{2^i} + \frac{1}{2^n}\right] \subset \left(\sum_{i=1}^{k_0} \frac{r_i}{2^i}, \sum_{i=1}^{k_0} \frac{r_i}{2^i} + \frac{1}{2^{k_0}}\right]$

En efecto:

$\sum_{i=1}^{k_0} \frac{s_i}{2^i} \leq \sum_{i=1}^{n} \frac{s_i}{2^i} < \sum_{i=1}^{n} \frac{s_i}{2^i} + \frac{1}{2^n} = \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \sum_{i=k_0+1}^{n} \frac{s_i}{2^i} + \frac{1}{2^n}$

$\leq \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \sum_{i=k_0+1}^{n} \frac{1}{2^i} + \frac{1}{2^n} = \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{\frac{1}{2^{k_0+1}} - \frac{1}{2^n}\frac{1}{2}}{\frac{1}{2}} + \frac{1}{2^n}$

$= \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{1}{2^{k_0}} - \frac{1}{2^n} + \frac{1}{2^n} = \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{1}{2^{k_0}}$

Y, haciendo un desarrollo similar:

$\sum_{i=1}^{k_0} \frac{r_i}{2^i} \leq \sum_{i=1}^{n} \frac{r_i}{2^i} < \sum_{i=1}^{n} \frac{r_i}{2^i} + \frac{1}{2^n} \leq \sum_{i=1}^{k_0} \frac{r_i}{2^i} + \frac{1}{2^{k_0}}$

Por lo tanto, $\left(\sum_{i=1}^{n} \frac{s_i}{2^i}, \sum_{i=1}^{n} \frac{s_i}{2^i} + \frac{1}{2^n}\right]$ y $\left(\sum_{i=1}^{n} \frac{r_i}{2^i}, \sum_{i=1}^{n} \frac{r_i}{2^i} + \frac{1}{2^n}\right]$ son ajenos.

Mostremos ahora que f es inyectiva:

Sean $(s_k)_{k\in\mathbb{N}}$ y $(r_k)_{k\in\mathbb{N}}$ dos elementos distintos de $\mathbb{B}$, $x = \sum_{k=1}^{\infty} \frac{s_k}{2^k}$ y $y = \sum_{k=1}^{\infty} \frac{r_k}{2^k}$, y definamos:

$k_0 = \text{mín}\{k \in \mathbb{N} : r_k \neq s_k\}$

Entonces:

$x \in \left(\sum_{i=1}^{k_0} \frac{s_i}{2^i}, \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{1}{2^{k_0}}\right]$

$y \in \left(\sum_{i=1}^{k_0} \frac{r_i}{2^i}, \sum_{i=1}^{k_0} \frac{r_i}{2^i} + \frac{1}{2^{k_0}}\right]$

Y, por lo anterior, $\left(\sum_{i=1}^{k_0} \frac{s_i}{2^i}, \sum_{i=1}^{k_0} \frac{s_i}{2^i} + \frac{1}{2^{k_0}}\right]$ y $\left(\sum_{i=1}^{k_0} \frac{r_i}{2^i}, \sum_{i=1}^{k_0} \frac{r_i}{2^i} + \frac{1}{2^{k_0}}\right]$ son ajenos.

Por lo tanto, $x \neq y$. Así que f es inyectiva.

Entonces la función f que asocia a cada sucesión $(s_k)_{k\in\mathbb{N}}$ en $\mathbb{B}$ el número $x = \sum_{k=1}^{\infty} \frac{s_k}{2^k}$ es una biyección de $\mathbb{B}$ en el intervalo $(0, 1]$.

Para cada $(r_1, r_2, \ldots, r_n) \in \mathbb{B}_n$, definamos:

$B_{r_1,r_2,\ldots,r_n} = \left\{(s_k)_{k\in\mathbb{N}} \in \mathbb{B} : \ s_k = r_k \text{ para cualquier } k \in \{1, 2, \ldots, n\}\right\}$

Entonces, $\bigcup_{\{(r_1,r_2,\ldots,r_n)\in\mathbb{B}_n\}} B_{r_1,r_2,\ldots,r_n} = \mathbb{B}$ y, por lo anterior:

$f\left(B_{r_1,r_2,\ldots,r_n}\right) \subset \left(\sum_{i=1}^{n} \frac{r_i}{2^i}, \sum_{i=1}^{n} \frac{r_i}{2^i} + \frac{1}{2^n}\right]$

Por lo tanto,

$\bigcup_{\{(r_1,r_2,\ldots,r_n)\in\mathbb{B}_n\}} \left(\sum_{i=1}^{n} \frac{r_i}{2^i}, \sum_{i=1}^{n} \frac{r_i}{2^i} + \frac{1}{2^n}\right] = (0, 1]$

Así que:

$f\left(B_{r_1,r_2,\ldots,r_n}\right) = \left(\sum_{i=1}^{n} \frac{r_i}{2^i}, \sum_{i=1}^{n} \frac{r_i}{2^i} + \frac{1}{2^n}\right]$

■

Corolario A.14. *1. Si* $(s_1, s_2, \ldots, s_n) \in \mathbb{B}_n$ *y* $x \in (0,1]$*, entonces,* $s_1, s_2, \ldots, s_n$ *son los primeros* n *términos del desarrollo de* x *en base* 2 *si y sólo si* $x \in \left(\sum_{i=1}^{n} \frac{s_i}{2^i}, \sum_{i=1}^{n} \frac{s_i}{2^i} + \frac{1}{2^n}\right]$.

2. Para cada $n \in \mathbb{N}$ *y cada intervalo* $\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]$*, donde* $k \in \{1, 2, \ldots, 2^n\}$*, existe un y sólo un elemento* $(s_1, s_2, \ldots, s_n) \in \mathbb{B}_n$ *tal que:*

$I_{(r_1,r_2,\ldots,r_n)} = \left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]$

3. La función $f_n : \mathbb{B}_n \to \left\{\frac{0}{2^n}, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, \frac{2^n-1}{2^n}\right\}$ *definida por:*

$f_n\left((s_1, s_2, \ldots, s_n)\right) = \sum_{k=1}^{n} \frac{s_k}{2^k}$

establece una relación uno a uno entre $\mathbb{B}_n$ *y el conjunto* $\left\{\frac{0}{2^n}, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, \frac{2^n-1}{2^n}\right\}$.

Referencias para la parte de historia

[1] Aristóteles; *Física*, Versión de Ute Schmidt Osmanczik, Biblioteca Scriptorum Graecorum et Romanorum Mexicana, Universidad Nacional Autónoma de México, México, 2005.

[2] Aristóteles; *Metafísica*, Traducción de Alex Sbantytown, Andrómeda Ediciones, Buenos Aires, 2003.

[3] Banach, S.; *Sur le problème de la mesure*, Fundamenta Mathematicae, 4, p. 7-33, 1923.

[4] Bernoulli, J.; *The Art of Conjecturing*, Traducción de Sylla, E. D., University Press, 2006. Traducción de *Ars Conjectandi*, Basileae, 1713.

[5] Borel, F. E. J. E.; *Sur quelques points de la Théorie des Fonctions*, C. R. Acad. Sci., t. 118, p. 340-342, 1894. Oeuvres de Émile Borel, Tome I, Centre National de la Recherche Scientifique, p. 235-237, 1972.

[6] Borel, F. E. J. E.; *Sur quelques points de la Théorie des Fonctions*, Thèse doctoral, Ann. Ec. Norm. Sup., 3em. série, t. 12, p. 9-55, 1895. Oeuvres de Émile Borel, Tome I, Centre National de la Recherche Scientifique, p. 239-285, 1972.

[7] Borel, F. E. J. E.; *Leçons sur la Théorie des Fonctions*, Gauthier-Villars, 1898.

[8] Borel, F. E. J. E.; *Remarques sur certains questions de Probabilité*, Bull. Soc. Math. Fr., T. 32, p. 123-128, 1904. Oeuvres de Émile Borel, Tome II, Centre National de la Recherche Scientifique, p. 985-990, 1972.

[9] Borel, F. E. J. E.; *Les probabilités dénombrables et leurs applications arithmétiques*, Rendiconti del Circolo Matematico di Palermo, T. 27, p. 247-270, 1909. Oeuvres de Émile Borel, Tome II, Centre National de la Recherche Scientifique, p. 1055-1079, 1972.

[10] Cantelli, F. P.; *Sulla probabilità comme limite della frequenza*, Rend. Acad. Lincei, Vol. 26, p. 39-45, 1917.

[11] Cantor, G. F. L. P.; *Ueber unendliche, lineare Punktmannichfaltigkeiten, Pt. 1*, Math. Ann., 15, p. 1-7, 1879.

[12] Cantor, G. F. L. P.; *Ueber unendliche, lineare Punktmannichfaltigkeiten, Pt. 2*, Math. Ann., 17, p. 355-358, 1880.

[13] Cantor, G. F. L. P.; *Ueber unendliche, lineare Punktmannichfaltigkeiten, Pt. 3*, Math. Ann., 20, p. 113-121, 1882.

[14] Cantor, G. F. L. P.; *Ueber unendliche, lineare Punktmannichfaltigkeiten, Pt. 4*, Math. Ann., 21, p. 51-58 y 545-591, 1883.

[15] Cantor, G. F. L. P.; *Ueber unendliche, lineare Punktmannichfaltigkeiten, Pt. 5*, Math. Ann., 23, p. 453-488, 1884.

[16] Carathéodory, C.; *Über das lineare Mass von Punktmengen - eine Verallgemeinerung des Längenbegriffs*, Nachrichten von der Königlichen Gesellchaft der Wiss zu Göttingen, p. 404-426, 1914.

[17] Cauchy, A. L.; *Résume des leçons données a l'Ëcole Royale Polytechnique sur le Calcul Infinitésimal*, Imprimerie Royale, 1823.

[18] Chebyshev, P. L.; *Des valeurs moyennes*, Matematicheskii Sbornik, 127, p. 1-9, 1867, también publicado en Liouville's Journal de Mathématiques Pures et Appliquées, 88, p.177-184, 1867.

[19] Chebyshev, P. L.; *Démonstration élémentaire d'une proposition générale de la théorie des probabilités.*

[20] Chebyshev, P. L.; *Sur deux théorèmes relatifs aux probabilités.*

[21] Daniell, P. J.; *A general form of integral*, Annals of Mathematics, Vol. 19, 1918.

[22] Daniell, P. J.; *Functions of limited variation in an infinite number of dimensions*, Annals of Mathematics, serie II, Vol. 21, p. 30-38, 1920.

[23] Daniell, P. J.; *Further properties of the general integral*, Annals of Mathematics, Serie II, Vol. 21, p. 203-220, 1920.

[24] Daniell, P. J.; *Integrals in an infinite number of dimensions*, Annals of Mathematics.

[25] de Finetti, B.; *Sui passaggi al limite nel Calcolo delle Probabilità*, (Reale) Istituto Lombardo de Science e Lettere, Rendiconti, Vol. 63, p. 155-166, 1930.

[26] de Finetti, B.; *A proposito dell'estensione del teorema delle probabilità totali alle classi numerabili*, (Reale) Istituto Lombardo de Science e Lettere, Rendiconti, Vol. 63, p. 901-905, 1930.

[27] de Finetti, B.; *Ancora sull'estensione alle classi numerabili del teorema delle probabilità totali*, (Reale) Istituto Lombardo de Science e Lettere, Rendiconti, Vol. 63, p. 1063-1069, 1930.

[28] de Moivre, A.; *Approximatio ad Summam Terminorum Binomii* $(a+b)^n$ *in Seriem expansi*, 1733.

[29] de Moivre, A.; *The doctrine of chances*, A. Millar, London, 1718 (third edition - 1756). Reimpreso por Chelsea, New York, 1967.

[30] du Bois-Reymond, P. D. G.; *Über die Integration der trigonometrischen Reihe*, Math. Ann., 22, p. 260-268, 1883.

[31] du Bois Reymond, P. D. G.; *Über die Integration der Reihen*, berlin Ak. Sber., p. 359-371, 1886.

[32] Fatou, P.; *Séries trigonométriques et séries de Taylor*, Acta. Mat., 30, 1906.

[33] Fermat, P. & Pascal, B.; *Correspondance-1654*, Oeuvres de Pascal, t. III, p. 369-430.

[34] Fischer, E.; *Sur la convergence en moyenne*, Comptes-rendus de l'Académie des Sciences de Paris, 144, 1907.

[35] Fourier, J. B. J.; *Théorie analytique de la chaleur*, Ed. Didot, 1822.

[36] Fréchet, M. J.; *Sur quelques points du Calcul Fonctionnel*, Rendiconti del Circolo Matematico di Palermo, 22, p. 1-74, 1906.

[37] Fréchet, M. R.; *Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait*, Bull. Soc. Mat. de France, 43, 1915.

[38] Fréchet, M. R.; *Des familles et fonctions additives d'ensembles abstraits*, Fundamenta Mathematicae, t. 4, 1923.

[39] Fréchet, M. R.; *Des familles et fonctions additives d'ensembles abstraits (Suite)*, Fundamenta Mathematicae, t. 5, 1924.

[40] Fréchet, M. R.; *Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements*, (Reale) Istituto Lombardo de Science e Lettere, Rendiconti, Vol. 63, p. 899-900, Milano, 1930.

[41] Fréchet, M. R.; *Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements (seconde note)*, (Reale) Istituto Lombardo de Science e Lettere, Rendiconti, Vol. 63, p. 1059-1062, 1930.

[42] Hald, A.; *A History of Probability and Statistics and Their Applications before 1750*, John Wiley, 1990.

[43] Hankel, H.; *Untersuchungen über die unendlich oft oszillierenden und unstetigen Functionen*, University of Tübingen, 1870, reproducido en Math. Ann., 20, 1882.

[44] Harnack, A.; *Die elemente der Differential und Integralrechnung*, B. G. Teubner, Leipzig, 1881.

[45] Harnack, A.; *Lehrbuch der Differential und Integralrechnung*, 2 Vols., B. G. Teubner, Leipzig, 1884-1885.

[46] Harnack, A.; *Über den Inhalt von Punktmengen*, Math. Ann. 25, p. 241-250, 1885.

[47] Hausdorff, F.; *Grundzüge der Mengenlehre*, Chelsea Publishing Company, 1914.

[48] Hilbert, D.; *Sur les problèmes futures des Mathématiques*, Comptes Rendus du Deuxième Congrès International des mathematiciens, Paris, p. 58-114, 1900.

[49] Huygens, C.; *Du calcul dans les jeux de hasard*, Oeuvres Complètes de Christiaan Huygens, Vol. XIV, Martinus Nijhoff, 1920. Traducción de De Ratiociniis in Aleae Ludo, 1657.

[50] Jordan, M. E. C;, *Cours d'Analyse de l'École Polytéchnique*, 3 Vols., Gauthier-Villars, 1882-1887. (Second edition, 1893-1896; Third edition, 1909).

[51] Khintchine, A.Ya.; *Sur la loi forte des grands nombres*, C. R. Ac. Sc. Paris, Vol. 186, p. 285-287, 1928.

[52] Kolmogorov, A. N.; *Sur la loi forte des grands nombres*, C. R. Ac. Sc. Paris, Vol. 191, p. 910-912, 1930.

[53] Kolmogorov, A. N.; *Foundations of the Theory of Probability*, Chelsea, 1950. Traducción de *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Erg Mat. 2, No. 3, 1933.

[54] Laplace, P. S.; *Théorie Analytique des Probabilités, Livre I. Calcul des fonctions génératrices*, Troisième edition, Courcier, Paris, 1820. Oeuvres complètes de Laplace, Tome septième, Gauthier-Villars, 1886.

[55] Laplace, P. S.; *Théorie Analytique des Probabilités, Livre II. Théorie générale des probabilités*, Troisième edition, Courcier, Paris, 1820. Oeuvres complètes de Laplace, Tome septième, Gauthier-Villars, 1886.

[56] Laplace, P. S.; *Essai philosophique sur les Probabilités* (1814), Gauthier-Villars, 1921.

[57] Lebesgue, H. L.; *Intégrale, longueur, aire*, Thèse doctoral, Ann. Math. Pur. Appl., 7 (3), p. 231-359, 1902.

[58] Lebesgue, H. L.; *Leçons sur l'intégration et la recherche des fonctions primitives*, Gauthier-Villars, 1904.

[59] Lebesgue, H. L.; *Sur la recherche des fonctions primitives par l'intégration*, R. Acc. Lincei Rend., (5), 1907.

[60] Lebesgue, H. L.; *L'intégration des fonctions discontinues*, Ann. Éc. Norm., 27 (3), 1910.

[61] Lévy, P. P.; *Les lois de probabilité dans les ensembles abstraits*, Revue de Métaphysique et Morale, 1924. Reproducido en Calcul des Probabilités, Gauthier Villars, 1925.

[62] Lévy, P. P.; *Calcul des Probabilités*, Gauthier Villars, Paris, 1925.

[63] Lyapunov, A. M.; *Sur une proposition de la Théorie des Probabilités*, Izv. Akad. Nauk., Ser. 5, 13, p. 359-386, 1900.

[64] Lyapunov, A. M.; *Nouvelle forme du théorème sur la limite des probabilités*, Notes Acad. Sci. Phys. Math. Sect., Ser. 8, 2, p. 1-24, 1901.

[65] Maistrov, L. E.; *Probability Theory - A historical sketch*, Academic press, 1974.

[66] Markov, A. A.; *The law of large numbers and the method of least squares*, Izd. Fiz. Mat. Ob.va Pri Kazan, Ser. 2, 8, p. 110-128, 1898.

[67] Markov, A. A.; *Extensión de la ley de los grandes números a variables dependientes*, Notices (Izvestiya) of the Physical Mathematical Society al Kazan University, Ser. 2, 15 (no.4), p. 155-156, 1907.

[68] Markov, A. A.; *Teorema Central del Límite para variables aleatorias dependientes*, 1908, 1910, 1911, 1912.

[69] Markov, A. A.; *Ischislenie Veroyatnostei (El Cálculo de Probabilidades)*, Moscow, 1913 (Cuarta edición, 1924).

[70] Michel, A.; *Constitution de la théorie moderne de l'intégration*, Librairie Philosophique J. Vrin, 1992.

[71] Moore, G. H., *Lebesgue's measure problem and Zermelo's Axiom of Choice: the mathematical effects of a philosophical dispute*, Ann. N. Y. Acad. Sci., 412, p. 129-154, 1983.

[72] Newman, J. R.; *Sigma, el mundo de las matemáticas, Vol. 3*, 1997.

[73] Nikodym, O.; *Sur une généralisation des intégrales de M. J. Radon*, Fundamenta Mathematicae, XV, p. 131-179, 1930.

[74] Peano, G., *Applicatione geometriche del Calcolo Infinitesimale*, Torino, 1887.

[75] Poincaré, J. H.; *Calcul des Probabilités*, Gauthier-Villars, París, 1896.

[76] Radon, J.; *Theorie une Anwendungen der absolut additiven Mengenfunktionen*, Sitzber der Math Naturwiss, Klasse der Kais, Akademie der Wiss, 122 (II.1), 1913.

[77] Riemann, G. F. B.; *Sur la possibilité de représenter une fonction par une série trigonométrique*, Mémoires de la Societé Royale des Sciences de Göttingue, t. XIII, 1867, traducción al francés reproducida en Oeuvres Mathématiques de Riemann, A. Blanchard, Paris, 1968.

[78] Riesz, F.; *Sur les systèmes orthogonaux de fonctions*, C. R. Ac. Sc., 144, 1907.

[79] Riesz, F.; *Sur une espèce de Géométrie analytique des systèmes de fonctions sommables*, C.R. 144, 1907.

[80] Riesz, F.; *Sur les opérations fonctionnelles linéaires*, C. R. Ac. Sc., 149, 1909.

[81] Riesz, F.; *Sur les suites des fonctions mesurables*, C. R. Ac. Sc., 148, 1909.

[82] Riesz, F.; *Sur quelques points de la théorie des fonctions sommables*, C. R. Ac. Sc., 154, 1912.

[83] Riesz, F.; *L'évolutionde la notion d'intégrale depuis Lebesgue*, Ann. Inst. Fourier, 1, 1949.

[84] Smith, H. J. S.; *On the integration of discontinuous functions*, London Math. Soc. Proc., 6, 1875.

[85] Solovay, R. M.; *A model of Set Theory in which every set of reals is Lebesgue measurable*, Ann. Math., 92, p. 1-56, 1970.

[86] Steinhaus, H. D.; *Les probabilités dénombrables et leur rapport à la Théorie de la Mesure*, Fundamenta Mathematicae, t. 4, p. 286-310, 1923.

[87] Stieltjes, T. J.; *Recherches sur les fractions continues*, Ann. Fac. Sc. Toul., t. VIII, 1894.

[88] Stolz, O.; *Über einen zu einer unendlichen Punktmenge gehörigen Grenzwerth*, Math. Ann., 23, p. 152-156, 1884.

[89] Stolz, O.; *Grundzüge der Differential und Integralrechnung*, 3 Vols., B. G. Teubner, Leipzig, 1893-99.

[90] Vitali, G.; *Sulle funzioni integrali*, Torino Acc. Sci. Atti, 40, 1904-1905.

[91] Vitali, G.; *Sul problema della misura dei gruppi di punti di una retta*, Tip. Bamberini et Parmeggiani, Bologna, 1905.

[92] Volterra, V., *Alcune osservazioni sulle funzioni punteggiate discontinue*, Giorn. Mat., 19, p. 76-86, 1881.

[93] Volterra, V.; *Sui principii del Calcolo Integrale*, Giorn. Mat., 19, p. 333-372, 1881.

[94] Von Mises, R.; *Grundlagen der Wahrscheinlichkeitsrechnung*, Math. Zeitsch, Vol. 5, p. 52-99, 1919.

[95] Von Mises, R.; *Mathematical Theory of Probability and Statistics*, 1919.

[96] Wiener, N.; *The mean of a functional of arbitrary elements*, Ann. of Math., (2) 22, p. 66-72, 1920.

[97] Wiener, N.; *The average of an analytic functional*, Proc. Nat. Acad. Sci. U.S.A., Vol. 7, No. 9, p. 253-260, 1921.

[98] Wiener, N.; *The average of an analytic functional and the Brownian Movement*, Proc. Nat. Acad. Sci. U.S.A., Vol. 7, No. 10, p. 294-298, 1921.

[99] Wiener, N.; *Differential space*, J. Math. and Physics, 2, p. 131-174, 1923.

[100] Wiener, N.; *Note on the series* $\sum\left(\pm\frac{1}{n}\right)$, Bull. Acad. Polon. Ser. A, 13, p. 83-90, 1923.

[101] Wiener, N.; *Un problème de probabilités dénombrables*, Bull. Soc. Math. France 11, p. 569-578, 1924.

[102] Wiener, N.; *The average value of a functional*, Proc. London Math. Soc., 22, p. 454-467, 1924.

Referencias para la formulación moderna

[1] Ash, R. B.; *Probability and Measure Theory*, second edition, Academic Press, 2000.
[2] Bachman, G. y Narici, L.; *Functional Analysis*, Academic Press, 1966.
[3] Bartle, R. G.; *Introducción al Análisis Matemático*, Limusa, 1982. Traducción de The elements of Real Analysis, 2a. edición, John Wiley, 1976.
[4] Billingsley, P.; *Probability and Measure*, John Wiley, 1979.
[5] Chung, K.L.; *A course in Probability Theory*, Second edition, Academic Press, 1974.
[6] García Álvarez, M. A.; *Introducción a la teoría de la probabilidad, primer curso*, FCE, 2005.
[7] García Álvarez, M. A.; *Introducción a la teoría de la probabilidad, segundo curso*, FCE, 2005.
[8] Royden, H. L.; *Real Analysis*, Second edition, Macmillan, 1968.

Índice