

TEORÍA DE LA MEDIDA

Sesión 19

El teorema de Kolmogorov

Introducción

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:

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.”

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**.

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

$$P(A \wedge B) = P(B | A) P(A)$$

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 $P(B | A)$ 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, \dots, A_n son eventos tales que ningún par de ellos puede ocurrir simultáneamente, entonces:

$$P(A_1 \vee A_2 \vee \dots \vee A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

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 fundamentado.

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:

“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 lo 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, 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.

Sin embargo, 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 artículo de Borel (*Les probabilités dénombrables et leurs applications arithmétiques*) el cual abrió una polémica acerca de las propiedades que debían pedirse a la función probabilidad en una formulación axiomática.

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, lo cual abría la pregunta acerca de la validez de esta propiedad para cualquier función de probabilidad. Si la respuesta era afirmativa se podría considerar a la función de probabilidad como una 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.

Por otra parte, en 1913, Johann Radon había ya desarrollado una teoría general de la medida en \mathbb{R}^n 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. 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 Fréchet 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 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. 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. Esta misma identificación la hizo Paul Pierre Lévy (1886-1971) en su libro *Calcul des Probabilités*, publicado en 1925, 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) y de Norbert Wiener (1894-1964).

Wiener trabajaba con funcionales lineales sobre espacios de funciones y seguía el método de Daniell para extender tales funcionales:

Con ese método, entre 1921 y 1923, Wiener construyó un modelo matemático para el movimiento browniano, para lo cual definió una medida de probabilidad σ -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.

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.

Sin embargo, la polémica sobre la propiedad de σ -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.

Finalmente, en el año 1933, Kolmogorov publicó un artículo titulado *Foundations of the Theory of Probability* 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.”

Kolmogorov estableció como modelo matemático de un fenómeno probabilístico una terna $(\Omega, \mathfrak{S}, P)$, donde Ω es un conjunto, \mathfrak{S} una σ -álgebra de subconjuntos Ω y P una medida de probabilidad definida sobre \mathfrak{S} .

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.

Finalmente, Kolmogorov, utilizando el método de Carathéodory, dio un método general, además de simple, para construir medidas de probabilidad en espacios de dimensión infinita:

Después del trabajo de Kolmogorov la aceptación de probabilidad como una medida fue unánime.

Regularidad de las medidas finitas sobre los borelianos de \mathbb{R}^n

Proposición 1. *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:*

$$d_k = \begin{cases} b_k + \delta & \text{si } b_k \in \mathbb{R} \\ b_k & \text{si } b_k = \infty \end{cases}$$

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:

$$F(x_1, \dots, x_n) = \mu((-\infty, x_1] \times \cdots \times (-\infty, x_n])$$

Para cualquier $\delta > 0$ definamos:

$$d_k^{(\delta)} = \begin{cases} b_k + \delta & \text{si } b_k \in \mathbb{R} \\ b_k & \text{si } b_k = \infty \end{cases}$$

$$R^{(\delta)} = (a_1, d_1^{(\delta)}] \times \cdots \times (a_n, d_n^{(\delta)}]$$

Se tiene entonces:

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \mu(R^{(\delta)}) \\ &= \lim_{\delta \rightarrow 0} \sum_{k=0}^n (-1)^k \sum \left\{ (x_1, \dots, x_n) \in S_{(a_1, d_1^{(\delta)}, \dots, a_n, d_n^{(\delta)})}^{(k)} \right\} F(x_1, \dots, x_n) \\ &= \sum_{k=0}^n (-1)^k \sum \left\{ (x_1, \dots, x_n) \in S_{(a_1, b_1, \dots, a_n, b_n)}^{(k)} \right\} F(x_1, \dots, x_n) = \mu(R) \end{aligned}$$

Dada $\varepsilon > 0$, existe entonces $\delta > 0$ tal que:

$$\mu(R^{(\delta)}) - \mu(R) < \varepsilon$$

Finalmente, como $R \subset R_\delta \subset R^{(\delta)}$, se tiene:

$$\mu(R) \leq \mu(R_\delta) \leq \mu(R^{(\delta)}) < \mu(R) + \varepsilon$$

■

Teorema 1. 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:

$$\mu(B) = \inf \{ \mu(O) : B \subset O \text{ y } O \text{ es un abierto de } \mathbb{R}^n \}.$$

Demostración

Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ la función de distribución finita definida por:

$$F(x_1, \dots, x_n) = \mu((-\infty, x_1] \times \dots \times (-\infty, x_n])$$

Sea \mathcal{I} la familia de rectángulos de la forma $(a_1, b_1] \times \dots \times (a_n, b_n]$ 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, \dots, 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:

$$\mu(B) = \inf \{ \sum_i \mu(A_i) : A_1, A_2, \dots \in \mathcal{A} \text{ y } B \subset \bigcup_i A_i \}$$

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, \dots , 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 $R^{(i,1)}, \dots, R^{(i,n_i)}$ son rectángulos en \mathcal{I} ajenos por parejas.

Para cada rectángulo $R^{(i,j)} = (a_1^{(i,j)}, b_1^{(i,j)}] \times \dots \times (a_n^{(i,j)}, b_n^{(i,j)}]$ sea $\delta_{ij} > 0$ tal que, si $R_{\delta_{ij}}^{(i,j)}$ es el rectángulo $(a_1^{(i,j)}, d_1^{(i,j)}) \times \dots \times (a_n^{(i,j)}, d_n^{(i,j)})$, 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 $\mu(R_{\delta_{ij}}^{(i,j)}) < \mu(R^{(i,j)}) + \frac{\varepsilon}{2^{i+j+1}}$.

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:

$$\begin{aligned} \mu(O_\varepsilon) &\leq \sum_i \sum_{j=1}^{n_i} \mu\left(R_{\delta_{ij}}^{(i,j)}\right) < \sum_i \left(\sum_{j=1}^{n_i} \mu\left(R^{(i,j)}\right) + \sum_{j=1}^{n_i} \frac{\varepsilon}{2^{i+j+1}} \right) \\ &\leq \sum_i \left(\mu(A_i) + \frac{\varepsilon}{2^{i+1}} \right) \leq \sum_i \mu(A_i) + \frac{\varepsilon}{2} \\ &< \mu(B) + \varepsilon \end{aligned}$$

■

Teorema 2. *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, \dots 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:

$$\mu(B) < \mu(K_N \cap B) + \frac{\varepsilon}{2} < \mu(K_N \cap O_N^c) + \varepsilon.$$

■

Conjuntos compactos

Definición 1. Sea (X, d) un espacio métrico

i) 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 $K \subset \bigcup_{\gamma \in T} G_\gamma$.

ii) 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 $K \subset \bigcup_{n \in T} G_n$.

iii) 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 .

Se puede demostrar que las propiedades *i*, *ii* y *iii*, enunciadas en la definición anterior, son equivalentes.

En la demostración del teorema de Kolmogorov vamos a utilizar el concepto de compacidad únicamente en \mathbb{R}^n ($n \in \mathbb{N}$). Para este caso, el teorema de Heine-Borel asegura que un conjunto es compacto si y sólo si es cerrado y acotado.

La propiedad que utilizaremos es la *iii* de la definición anterior. Es por eso que únicamente vamos a demostrar que *i* implica *iii*.

Teorema 1. Si K es un conjunto compacto de \mathbb{R}^n , entonces es secuencialmente compacto.

Demostración

Sea $(x_i)_{i \in \mathbb{N}}$ una sucesión cualquiera de elementos de K ,

Como K es acotado, existe una celda $I_1^{(1)} \times I_2^{(1)} \times \dots \times I_n^{(1)}$ formada por intervalos cerrados finitos que lo contiene, a la cual llamaremos C_1 . Podemos tomarla de tal forma que los intervalos I_1, I_2, \dots, I_n tengan la misma longitud, la cual denotaremos por L .

Vamos a construir, inductivamente, una sucesión $(C_m)_{m \in \{2, 3, \dots\}}$ de celdas, formadas por intervalos cerrados finitos, tales que, para cualquier $m \in \{2, 3, \dots\}$:

i) $C_m \subset C_{m-1}$.

ii) C_m contiene una infinidad de elementos de la sucesión $(x_i)_{i \in \mathbb{N}}$.

iii) Si $C_m = I_1^{(m)} \times I_2^{(m)} \times \dots \times I_n^{(m)}$, entonces, para cualquier $j \in \{1, 2, \dots, n\}$, $l(I_j^{(m)}) = \frac{1}{2^{(m-2)}} L$.

Definiendo $C_2 = C_1$, C_2 cumple con las condiciones *i* *ii* y *iii*.

Tomemos ahora $k \in \{2, 3, \dots\}$ y supongamos que tenemos definida una celda $C_k = [a_1^{(k)}, b_1^{(k)}] \times [a_2^{(k)}, b_2^{(k)}] \times \dots \times [a_n^{(k)}, b_n^{(k)}]$ satisfaciendo las propiedades i ii y iii.

Para cada $j \in \{1, 2, \dots, n\}$, denotemos por $c_j^{(k)}$ al punto medio del intervalo $[a_j^{(k)}, b_j^{(k)}]$.

De esta forma, en cada coordenada j tenemos los intervalos $[a_j^{(k)}, c_j^{(k)}]$ y $[c_j^{(k)}, b_j^{(k)}]$.

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 $C = I_1 \times I_2 \times \dots \times I_n$ es cualquiera de esas celdas, se tiene $C \subset C_k$ y, para cualquier $j \in \{1, 2, \dots, n\}$, se tiene:

$$l(I_j) = \frac{1}{2}l\left([a_j^{(k)}, b_j^{(k)}]\right) = \frac{1}{2} \frac{1}{2^{(k-2)}}L = \frac{1}{2^{(k-1)}}L$$

Sabemos que C_k contiene una infinidad de elementos de la sucesión $(x_i)_{i \in \mathbb{N}}$, así que, por lo menos una de las 2^n celdas que formamos, llamémosla C , contiene una infinidad de elementos de la sucesión $(x_i)_{i \in \mathbb{N}}$, porque si para cualquiera de las 2^n celdas se tuviera la propiedad contraria, C_k contendría únicamente un número finito de elementos de la sucesión $(x_i)_{i \in \mathbb{N}}$.

Definamos entonces C_{k+1} como una cualquiera de esas celdas C , entre las 2^n celdas que formamos, con la propiedad de que contiene una infinidad de elementos de la sucesión $(x_i)_{i \in \mathbb{N}}$.

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 $m \in \{2, 3, \dots\}$, queda definida cada una de las celdas C_m satisfaciendo las propiedades i, ii y iii.

Para cada $m \in \{2, 3, \dots\}$, 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 $\bigcap_{m=1}^{\infty} C_m \neq \emptyset$. Esta intersección es un conjunto formado por un único punto, ya que si $x = (x_1, x_2, \dots, x_n)$ y $y = (y_1, y_2, \dots, y_n)$ pertenecen a esa intersección, entonces, para cada $j \in \{1, 2, \dots, n\}$ y cualquier $m \in \{2, 3, \dots\}$, 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 $m \in \{2, 3, \dots\}$ y, entonces, $x_j = y_j$.

Sea $z = (z_1, z_2, \dots, z_n)$ el único punto en la intersección $\bigcap_{m=1}^{\infty} C_m$.

Tomemos cualquier elemento de la sucesión $(x_n)_{n \in \mathbb{N}}$, digamos $x_{i_1} = (x_1^{(i_1)}, x_2^{(i_1)}, \dots, x_n^{(i_1)})$. Como, para cada $m \in \{2, 3, \dots\}$, C_m contiene una infinidad de elementos de la sucesión $(x_i)_{i \in \mathbb{N}}$, podemos elegir, para cada $m \in \{2, 3, \dots\}$, un elemento $x_{i_m} = (x_1^{(i_m)}, x_2^{(i_m)}, \dots, x_n^{(i_m)})$

de la sucesión $(x_i)_{i \in \mathbb{N}}$ de tal forma que $x_{i_m} \in C_m$ y $i_m > i_{m-1}$. De esta forma, la sucesión $(x_{i_m})_{m \in \mathbb{N}}$ es una subsucesión de $(x_i)_{i \in \mathbb{N}}$.

Para cada $m \in \{2, 3, \dots\}$, tanto z como x_{i_m} pertenecen a la celda C_m , así que, para cualquier $j \in \{1, 2, \dots, n\}$, se tiene:

$$\left| x_j^{(i_m)} - z_j \right| \leq L^{(m)}$$

Por lo tanto:

$$\begin{aligned} d(x_{i_m}, z) &= \sqrt{\left(x_1^{(i_m)} - z_1\right)^2 + \left(x_2^{(i_m)} - z_2\right)^2 + \dots + \left(x_n^{(i_m)} - z_n\right)^2} \\ &\leq L^{(m)} \sqrt{n} = \frac{1}{2^{(m-2)}} L \sqrt{n} \end{aligned}$$

Así que la sucesión $(x_{i_m})_{m \in \mathbb{N}}$ converge a z .

Además, $x_{i_m} \in K$ para cualquier $m \in \mathbb{N}$, así que, como K es cerrado, $z = \lim_{m \rightarrow \infty} x_{i_m} \in K$.

Por lo tanto, hemos encontrado una subsucesión de $(x_i)_{i \in \mathbb{N}}$ la cual converge a algún elemento de K , lo cual prueba el resultado. ■

Teorema de Kolmogorov

El teorema de Kolmogorov muestra que la propiedad de σ -aditividad de una función de probabilidad es consistente en el sentido de que si se asume como válida para el caso finito, entonces se puede extender al caso infinito.

Se puede enunciar de la manera siguiente:

Teorema 3 (Teorema de Kolmogorov). *Sea Γ un conjunto infinito y supongamos que para cada subconjunto finito u de Γ , con n elementos, se tiene una función de distribución en n variables $F_u : \mathbb{R}^n \rightarrow \mathbb{R}$ de tal forma que si v y w son dos subconjuntos finitos de Γ tales que $v \subset w$, entonces la función de distribución F_v coincide con la distribución marginal que se obtiene de F_w restringiéndola a las coordenadas en u . Entonces, existe un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ y una familia de variables aleatorias reales $\{X_t : t \in \Gamma\}$ 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 .*

Obsérvese que, al inicio, no se tienen variables aleatorias ya que esto asumiría que se tiene un espacio de probabilidad en el cual están definidas y es precisamente ese espacio de probabilidad el que se construye en la demostración del teorema.

Por ejemplo, en el caso del proceso de Wiener, el cual modela el movimiento, en una dimensión, de un grano de polen que se coloca sobre agua (movimiento browniano), lo que se tiene de inicio es lo siguiente:

Imaginando que el grano de polen se mueve en un plano cartesiano infinito, considerando la proyección del movimiento sobre uno de los ejes, digamos el horizontal, se tiene una partícula que se mueve sobre una línea recta, así que, en cada tiempo $t \geq 0$, la partícula se encuentra en una posición que denotaremos por W_t . Esta posición es aleatoria ya que el movimiento del grano de polen se debe a los choques que recibe de las moléculas de agua. Lo que se busca entonces es un espacio de probabilidad en el cual se pueda definir una familia de variables aleatorias reales $\{W_t : t \in [0, \infty)\}$ de tal manera que W_t represente la posición de la partícula en el tiempo t . Para esto, se parte de las propiedades que se observan en un movimiento browniano, o mejor dicho, de propiedades que se aproximan a las observaciones. Estas propiedades consisten en lo siguiente:

1. El grano de polen no se mueve a saltos, sino de una manera continua; así que, cualquiera que sea su movimiento, la función $t \rightarrow W_t$ es continua.
2. El punto de inicio del movimiento se toma como el origen de la línea recta sobre el cual se mueve la partícula. Así que $W_0 = 0$.
3. El movimiento que sigue la partícula a partir de su posición en un tiempo $t > 0$, es independiente de como haya sido en el intervalo $[0, t]$. Así que, Si $t_1 < t_2 < \dots < t_n$ son números reales positivos, entonces las variables aleatorias $W_{t_1} - W_0, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ son independientes.
4. Si $0 \leq s < t$, la distribución de $W_t - W_s$ es normal con parámetros $\mu = 0$ y $\sigma^2 = t - s$.

Si $t_1 < t_2 < \dots < t_n$ son números reales positivos, sabiendo que las variables aleatorias $W_{t_1} - W_0, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ son independientes, podemos encontrar la función de densidad conjunta de $W_{t_1}, W_{t_2}, \dots, W_{t_n}$ considerando la transformación $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ dada por:

$$y_k = x_1 + x_2 + \dots + x_k, \text{ para } k \in \{1, \dots, n\}.$$

cuya inversa, ϕ , está dada por:

$$x_1 = y_1$$

$$x_k = y_k - y_{k-1}, \text{ para } k \in \{2, \dots, n\}.$$

En efecto, se tiene:

$$(W_{t_1}, \dots, W_{t_n}) = \varphi(W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}})$$

Así que:

$$\begin{aligned}
& f_{W_{t_1}, \dots, W_{t_n}}(y_1, \dots, y_n) \\
&= |J_\phi(y_1, \dots, y_n)| f_{W_{t_1}, W_{t_2-W_{t_1}}, \dots, W_{t_n-W_{t_{n-1}}}}(\phi(y_1, \dots, y_n)) \\
&= \frac{1}{\sqrt{(2\pi)^n t_1(t_2-t_1)\dots(t_n-t_{n-1})}} \exp\left\{-\frac{1}{2t_1}y_1^2\right\} \exp\left\{-\frac{1}{2(t_2-t_1)}(y_2-y_1)^2\right\} \cdots \exp\left\{-\frac{1}{2(t_n-t_{n-1})}(y_n-y_{n-1})^2\right\} \\
&= \frac{1}{\sqrt{(2\pi)^n t_1(t_2-t_1)\dots(t_n-t_{n-1})}} \exp\left\{-\frac{1}{2}\left[\frac{1}{t_1}y_1^2 + \frac{1}{t_2-t_1}(y_2-y_1)^2 + \cdots + \frac{1}{t_n-t_{n-1}}(y_n-y_{n-1})^2\right]\right\}
\end{aligned}$$

Algo similar se puede hacer cuando $t_1 = 0$.

En otras palabras, para cada subconjunto finito $\{t_1, t_2, \dots, t_n\}$ de números reales no negativos, se tiene la función de densidad conjunta, y, por lo tanto, la función de distribución conjunta, de $W_{t_1}, W_{t_2}, \dots, W_{t_n}$.

De lo que se trata entonces es de que, partiendo del conocimiento de esas distribuciones conjuntas, se pueda construir un espacio de probabilidad en el cual se pueda definir toda la familia de variables aleatorias $\{W_t : t \in [0, \infty)\}$ de tal manera que para cada subconjunto finito de números reales no negativos $\{t_1, t_2, \dots, t_n\}$ la función de distribución conjunta de $W_{t_1}, W_{t_2}, \dots, W_{t_n}$ sea la que se tenía al inicio.

Regresando a la formulación general del teorema de Kolmogorov, la idea de la demostración es la siguiente:

Para cada subconjunto finito de 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}^Γ . Para cada subconjunto finito u de Γ se expresa \mathbb{R}^Γ 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}^Γ 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}^Γ 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 y los conjuntos de medida exterior cero.

Recordemos que si \mathbb{F} es un conjunto cualquiera y \mathcal{A} un álgebra de subconjuntos de \mathbb{F} , una quasi medida es una función no negativa $\mu : \mathcal{A} \mapsto \overline{\mathbb{R}}$ finitamente aditiva, σ -subaditiva y tal que $\mu(\emptyset) = 0$.

Además, si \mathbb{F} es un conjunto cualquiera, \mathcal{A} un álgebra de subconjuntos de \mathbb{F} y $\mu : \mathcal{A} \rightarrow \mathbb{R}$ 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 $\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \mu(A_n)$.

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 $\mu(\bigcap_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \mu(A_n)$.

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 $\lim_{n \rightarrow \infty} \mu(A_n) = 0$.

Finalmente, recordemos el teorema de extensión de Carathéodory:

Si \mathbb{F} es un conjunto cualquiera, \mathcal{A} un álgebra de subconjuntos de \mathbb{F} y $\mu_0 : \mathcal{A} \rightarrow \overline{\mathbb{R}}$ una quasi medida, entonces existe una medida $\mu : \mathfrak{S} \rightarrow \overline{\mathbb{R}}$ tal que $\mu(A) = \mu_0(A)$ para cualquier $A \in \mathcal{A}$, donde \mathfrak{S} es una σ -álgebra que contiene a $\sigma(\mathcal{A})$ y a los conjuntos de medida exterior cero.

La medida μ se construye definiendo primero una medida exterior definida sobre todos los subconjuntos de \mathbb{F} :

$$\mu_e(A) = \inf \left\{ \sum_j \mu_0(A_j) : A_1, A_2, \dots \text{ es cubierta de } A \right\}$$

Después se definen los conjuntos medibles como aquellos conjuntos $E \subset \mathbb{F}$ para los cuales se cumple la siguiente relación:

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

para cualquier conjunto $A \subset \mathbb{F}$.

La medida $\mu(E)$ de un conjunto medible E se define entonces como la medida exterior de E .

Demostración del teorema de Kolmogorov

Paso 1.

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 $t \in \Gamma$.

Paso 2.

Sabemos que, para cada $u \in U$, con n elementos, se tiene una función de distribución en n variables $F_u : \mathbb{R}^u \rightarrow \mathbb{R}$.

Esta función de distribución F_u genera una medida de probabilidad P_u definida sobre los conjuntos borelianos de \mathbb{R}^u tal que:

$$P_u((-\infty, u_1] \times \cdots \times (-\infty, u_n]) = F_u(u_1, \dots, u_n)$$

para cualquier $(u_1, \dots, u_n) \in \mathbb{R}^u$.

Paso 3.

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.

Paso 4.

Sabemos que si $u, v \in U$ y $u \subset v$, entonces la función de distribución F_u coincide con la distribución marginal que se obtiene de F_v restringiéndola a las coordenadas de u ; es decir, si $u = \{u_1, \dots, u_n\}$ y $v = \{u_1, \dots, u_n, v_1, \dots, v_m\}$, entonces:

$$F_u(u_1, \dots, u_n) = \lim_{(v_1, \dots, v_m) \rightarrow (\infty, \dots, \infty)} F_v(u_1, \dots, u_n, v_1, \dots, v_m)$$

Así que:

$$P_u((-\infty, u_1] \times \cdots \times (-\infty, u_n]) = P_v((-\infty, u_1] \times \cdots \times (-\infty, u_n] \times \mathbb{R} \times \cdots \times \mathbb{R})$$

Pero:

$$(-\infty, u_1] \times \cdots \times (-\infty, u_n] \times \mathbb{R} \times \cdots \times \mathbb{R} = \Pi_{vu}^{-1} ((-\infty, u_1] \times \cdots \times (-\infty, u_n])$$

Por lo tanto:

$$P_u ((-\infty, u_1] \times \cdots \times (-\infty, u_n]) = P_v (\Pi_{vu}^{-1} ((-\infty, u_1] \times \cdots \times (-\infty, u_n]))$$

Como los conjuntos de la forma $(-\infty, u_1] \times \cdots \times (-\infty, u_n]$ generan la σ -álgebra de Borel en \mathbb{R}^u , se tiene entonces:

$$P_u(B_u) = P_v (\Pi_{vu}^{-1} (B_u))$$

para cualquier conjunto boreliano B_u de \mathbb{R}^u .

Paso 5.

Para cada $u \in U$, denotemos por \mathcal{B}_u a la σ -álgebra de los conjuntos borelianos de \mathbb{R}^u y definamos:

$$\mathfrak{S}_0 = \{\Pi_u^{-1} (B_u) : u \in U \text{ y } B_u \in \mathcal{B}_u\}$$

Cada elemento de \mathfrak{S}_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 $\mathbb{R}^{\Gamma-u}$.

Paso 6.

Demostremos que \mathfrak{S}_0 es un álgebra de subconjuntos de Ω :

Obviamente, $\Omega \in \mathfrak{S}_0$ y si $E \in \mathfrak{S}_0$ entonces $E^c \in \mathfrak{S}_0$. Por otra parte, si $E = \Pi_u^{-1} (A_u) \in \mathfrak{S}_0$ y $F = \Pi_v^{-1} (B_v) \in \mathfrak{S}_0$, sea $w = u \cup v$, $A_w = \Pi_{wu}^{-1} (A_u)$ y $B_w = \Pi_{wv}^{-1} (B_v)$, entonces:

$$E \cup F = \Pi_w^{-1} (A_w \cup B_w) \in \mathfrak{S}_0.$$

Por lo tanto, \mathfrak{S}_0 es un álgebra de subconjuntos de Ω .

Paso 7.

Definamos $P : \mathfrak{S}_0 \mapsto [0, 1]$ de la siguiente manera:

$$P (\Pi_u^{-1} (B_u)) = P_u (B_u)$$

Observemos en primer lugar que P está bien definida. En efecto, supongamos que $\Pi_u^{-1}(B_u) = \Pi_v^{-1}(A_v)$, entonces, definiendo $w = u \cup v$, se tiene:

$$\Pi_u^{-1}(B_u) = \Pi_w^{-1}(\Pi_{wu}^{-1}(B_u))$$

$$\Pi_v^{-1}(A_v) = \Pi_w^{-1}(\Pi_{wv}^{-1}(A_v))$$

Así que:

$$\Pi_{wu}^{-1}(B_u) = \Pi_{wv}^{-1}(A_v)$$

Por lo tanto:

$$P_u(B_u) = P_w(\Pi_{wu}^{-1}(B_u)) = P_w(\Pi_{wv}^{-1}(A_v)) = P_v(A_v)$$

Paso 8.

$\Omega = \Pi_u^{-1}(\mathbb{R}^u)$ para cualquier $u \in U$.

Así que:

$$P(\Omega) = P(\Pi_u^{-1}(\mathbb{R}^u)) = P_u(\mathbb{R}^u) = 1$$

Paso 9.

Mostremos que P es finitamente aditiva:

En efecto, Si $E = \Pi_u^{-1}(A_u)$ y $F = \Pi_v^{-1}(B_v)$ son elementos de \mathfrak{S}_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 $E \cup F = \Pi_w^{-1}(A_w \cup B_w)$, así que:

$$\begin{aligned} P(E \cup F) &= P_w(A_w \cup B_w) = P_w(A_w) + P_w(B_w) \\ &= P_u(A_u) + P_v(B_v) = P(E) + P(F) \end{aligned}$$

Paso 10.

Mostremos ahora que P es σ -subaditiva:

Para esto, basta con demostrar que si tenemos una sucesión decreciente $(E_i)_{i \in \mathbb{N}}$, de elementos de \mathfrak{S}_0 , tal que $\bigcap_{i=1}^{\infty} E_i = \emptyset$, entonces $\lim_{i \rightarrow \infty} P(E_i) = 0$.

Para cada $i \in \mathbb{N}$, sea $v_i \in U$ y $A_i \in \mathcal{B}_{v_i}$ tales que $E_i = \Pi_{v_i}^{-1}(A_i)$, y definamos $u_i = \bigcup_{j=1}^i v_j$ y $B_i = \Pi_{u_i v_i}^{-1}(A_i)$. Entonces $B_i \in \mathcal{B}_{u_i}$, $E_i = \Pi_{u_i}^{-1}(B_i)$ y la sucesión de conjuntos $(u_n)_{n \in \mathbb{N}}$ es creciente.

Supongamos que $\varepsilon = \lim_{i \rightarrow \infty} P(E_i) > 0$.

Por el teorema 2, para cada $i \in \mathbb{N}$, $P_{u_i}(B_i)$ 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:

$$P_{u_i}(B_i) - P_{u_i}(K_i) < \frac{\varepsilon}{2^{i+1}}$$

Sea $F_i = \Pi_{u_i}^{-1}(K_i)$.

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:

$$\begin{aligned} P(E_j) - P(H_j) &= P(E_j) - P\left(\bigcap_{i=1}^j F_i\right) = P\left(\bigcup_{i=1}^j (E_j - F_i)\right) \\ &\leq \sum_{i=1}^j P(E_j - F_i) \leq \sum_{i=1}^j P(E_i - F_i) = \sum_{i=1}^j [P(E_i) - P(F_i)] \\ &= \sum_{i=1}^j [P_{u_i}(B_i) - P_{u_i}(K_i)] < \sum_{i=1}^j \frac{\varepsilon}{2^{i+1}} < \frac{\varepsilon}{2} \end{aligned}$$

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 $i \in \{1, \dots, j\}$. Por lo tanto, $x_{u_i}^{(j)} = \Pi_{u_i} x^{(j)} \in K_i$ para $i \in \{1, \dots, j\}$.

(Obsérvese que, para cualesquiera $i, j \in \mathbb{N}$, $x_{u_i}^{(j)}$ tiene como coordenadas las coordenadas de $x^{(j)}$ que corresponden a los elementos de u_i . Recordemos, además, que $u_1 \subset u_2 \subset u_3 \subset \dots$)

Visto de otra manera, fijando $i \in \mathbb{N}$ se tiene $x_{u_i}^{(j)} \in K_i$ para cualquier $j \in \{i, i+1, \dots\}$.

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 $m(1,1) > 1$.

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 $m(2,1) > 2$.

Además:

$$\begin{aligned}\Pi_{u_2 u_1}(x_{u_2}^{(m(2,j))}) &= x_{u_1}^{(m(2,j))} \\ \lim_{j \rightarrow \infty} x_{u_1}^{(m(2,j))} &= \lim_{j \rightarrow \infty} x_{u_1}^{(m(1,j))}\end{aligned}$$

Así que, si $x_{u_1} = \lim_{j \rightarrow \infty} x_{u_1}^{(m(1,j))}$ y $x_{u_2} = \lim_{j \rightarrow \infty} x_{u_2}^{(m(2,j))}$, 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 $x_{u_i} = \lim_{j \rightarrow \infty} x_{u_i}^{(m(i,j))}$, 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 $\Pi_{u_j u_i}(x_{u_j}) = x_{u_i}$.

Definamos la función $y : \cup_{i=1}^{\infty} u_i \rightarrow \mathbb{R}$ de la siguiente manera:

$$y(t) = x_{u_i}(t) \text{ si } t \in u_i$$

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 $y \in \mathbb{R}^{\cup_{i=1}^{\infty} u_i}$, 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 $\omega_0 \in \Omega$ de la siguiente manera:

$$\omega_0(t) = \begin{cases} y(t) & \text{si } t \in \cup_{i=1}^{\infty} u_i \\ 0 & \text{en otro caso} \end{cases}$$

Entonces $\Pi_{u_i}(\omega_0) = x_{u_i} \in K_i$ para cualquier $i \in \mathbb{N}$, así que $\omega_0 \in \Pi_{u_i}^{-1}(K_i) = F_i$ para cualquier $i \in \mathbb{N}$, es decir:

$$\omega_0 \in \bigcap_{i=1}^{\infty} F_i$$

Así que $\bigcap_{i=1}^{\infty} F_i \neq \emptyset$, lo cual establece la contradicción mencionada.

Por lo tanto $\lim_{i \rightarrow \infty} P(E_i) = 0$, así que P es σ -subaditiva.

De acuerdo con el teorema de extensión de Carathéodory, existe una única medida de probabilidad P definida sobre $\mathfrak{S} = \sigma(\mathfrak{S}_0)$ 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:

$$[X_t \leq x] = \Pi_{\{t\}}^{-1}((-\infty, x]) \in \mathfrak{S}_0$$

para cualquier $t \in \Gamma$ y $x \in \mathbb{R}$.

Así que X_t es \mathfrak{S} -medible para cualquier $t \in \Gamma$.

Además, para cualquier $u = \{t_1, \dots, t_n\} \in U$ y $(x_1, \dots, x_n) \in \mathbb{R}^u$, se tiene:

$$\begin{aligned} P[X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n] &= P(\Pi_u^{-1}((-\infty, x_1] \times \dots \times (-\infty, x_n])) \\ &= P_u((-\infty, x_1] \times \dots \times (-\infty, x_n]) = F_u(x_1, \dots, x_n). \end{aligned}$$

Así que F_u es la función de distribución conjunta de X_{t_1}, \dots, X_{t_n} .

■