

¿QUÉ ES EL CÁLCULO ESTOCÁSTICO?

Versión simplificada

Miguel Ángel García Álvarez

1. INTRODUCCIÓN

El Cálculo Estocástico es en realidad únicamente Cálculo Integral Estocástico, por lo menos en lo que se refiere a la teoría que quedó prácticamente concluída hacia finales de los años 70's del siglo pasado y que se puede denominar como Cálculo Estocástico Clásico. Después esa teoría se fue ampliando y se desarrolló en varias direcciones. Ahora bien, al igual que ocurrió con el Cálculo Diferencial e Integral que conocemos, la teoría de Integración Estocástica se desarrolló para poderla aplicar en la solución de cierto tipo de problemas de interés. De manera específica, el objetivo es poder resolver las llamadas ecuaciones diferenciales estocásticas, las cuales también son en realidad ecuaciones integrales estocásticas. Como veremos más adelante, en el Cálculo Estocástico Clásico no hay derivadas; es decir, no hay un Cálculo Diferencial Estocástico ya que, en general, se trabaja con funciones que no son diferenciables. A su vez, el resolver ecuaciones diferenciales estocásticas tiene por objetivo el poder construir procesos estocásticos que modelen fenómenos de interés en diversas áreas. Una de las aplicaciones más desarrolladas es la referente a las Finanzas, pero actualmente las aplicaciones están más diversificadas. Por otra parte, si bien, como lo mencioné antes, la teoría del Cálculo Estocástico Clásico quedó bien establecida en los 70's del siglo pasado, no fue así en lo que se refiere a las ecuaciones diferenciales estocásticas, tema que actualmente aún sigue en desarrollo.

Para entender el concepto de integral estocástica es muy útil tener una idea clara de cómo se fue desarrollando la teoría de integración tal como se enseña actualmente en los cursos de Cálculo (integral de Riemann) y en los de Análisis Matemático (integral de Lebesgue). Esto no únicamente por curiosidad, sino porque las investigaciones alrededor del concepto de integral y sus propiedades condujeron a lo que ahora se llama Teoría de la Medida, la cual es una de las herramientas fundamentales que se utilizan en el Cálculo Estocástico.

2. TEORÍA DE LA MEDIDA

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 geoméricamente 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 geoméricamente la interpretaban 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.

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

Georg Friedrich Bernhard Riemann, en un artículo titulado *Sur la possibilité de représenter une fonction par une série trigonométrique*, 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, \dots, 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, mediante los cuales aclaró algunas confusiones que había en ese momento con relación a la integrabilidad de una función; en particular, exhibió una función integrable, definida en un intervalo cerrado, para la cual el conjunto de sus discontinuidades es denso en el intervalo. El ejemplo de Riemann echó abajo una idea que se tenía según la cual para que una función sea integrable se requiere que el conjunto de sus discontinuidades sea topológicamente pequeño.

Definición 1. Sea $f : [a, b] \rightarrow \mathbb{R}$ una función acotada. La diferencia:

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

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

Definición 2. Sea $f : [a, b] \rightarrow \mathbb{R}$ una función acotada, $x \in (a, b)$ e $(I_n)_{n \in \mathbb{N}}$ 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 $\lim_{n \rightarrow \infty} O_n$ 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** .

Después de la publicación del artículo de Riemann, las investigaciones se orientaron hacia la búsqueda de una caracterización de las funciones integrables, basada en el tipo de conjunto

que forman sus discontinuidades. Fue así que se llegó a un nuevo concepto, el de contenido cero.

Definición 3. *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 .*

Una vez definido este concepto se pudo establecer con toda claridad la condición para que una función sea integrable. Axel Harnack demostró en 1881 que **una función es integrable si y sólo si, para cualquier $\sigma > 0$, el conjunto de puntos donde la oscilación de la función es mayor que σ tiene contenido cero.**

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, sobre todo, por Camille Jordan 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 ese periodo los conceptos de integral superior e inferior de una función.

Definición 4. *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 las longitudes de los subintervalos de P que contienen puntos de A y $\underline{S}(P, A)$ la suma de las longitudes 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:*

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

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

*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.

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 tiene el siguiente resultado:

Sea $f : [a, b] \rightarrow \mathbb{R}$ 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 $\int_a^b f(x)dx = c(E)$.

En 1894-1895, Émile Borel dio las bases para un nuevo avance al introducir el concepto de medida cero:

Definición 5. *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 .*

Además de introducir el concepto de medida cero, al resolver el problema que lo motivó para dar esa definición, 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 e $(I_j)_{j \in \mathbb{N}}$ 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.

Para demostrar el resultado enunciado, Borel demostró que existe una colección finita de los intervalos I_j cuya unión también contiene a I , resultado que más tarde llevó al concepto de compacidad.

Más adelante, en un libro publicado en 1898, Borel retomó el concepto de conjunto de medida cero para desarrollar una teoría de la medida. 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. Las propiedades esenciales que planteó Borel son las siguientes:

1. La medida de la unión de una colección numerable de conjuntos ajenos es igual a la suma de sus medidas.
2. 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 $m(B) - m(A)$.

3. La medida de un conjunto nunca es negativa.

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 Lebesgue en 1902.

Una función acotada $f : [a, b] \mapsto \mathbb{R}$ 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*. Más tarde la expuso en su libro *Leçons sur l'intégration et la recherche des fonctions primitives*.

Mostró Lebesgue que para dar una definición constructiva de la integral de cualquier función acotada, basta con hacerlo para las funciones que únicamente toman los 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:

1. Si E es un conjunto acotado y $a \in \mathbb{R}$, entonces $m(E + a) = m(E)$.
2. Si E_1, E_2, \dots es una familia finita o infinita numerable de conjuntos, ajenos por parejas y contenidos en un conjunto acotado, entonces $m(\bigcup_n E_n) = \sum_n m(E_n)$.
3. $m([0, 1]) = 1$.

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

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

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

Si E es un conjunto acotado e I_1, I_2, \dots es una colección finita o infinita numerable de intervalos abiertos tales que $E \subset \bigcup_n I_n$, 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:

$$m_e(E) = \inf \left\{ \sum_n l(I_n) : I_1, I_2, \dots \text{ son intervalos abiertos y } E \subset \bigcup_n I_n \right\}$$

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 definió entonces como la **medida interior** de E y la denotó por $m_i(E)$.

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

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.

Lo mismo ocurre con la teoría de integración que desarrolló Lebesgue a partir de su concepto de medida. La integral de Lebesgue resulta ser más general que la de Riemann: toda función Riemann-integrable es Lebesgue-integrable, pero hay muchísimas funciones que son Lebesgue-integrables, pero no Riemann-integrables. Además, la integral definida por Lebesgue tiene propiedades que no las tiene la integral de Riemann y eso la hace mucho más manejable.

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*, cuyo desarrollo 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 $\{x_0, x_1, \dots, x_n\}$ del intervalo $[a, b]$, tomaba un punto ξ_i en cada subintervalo $[x_{i-1}, x_i]$, consideraba la suma $\sum_{i=1}^n \xi_i (\Phi(x_i) - \Phi(x_{i-1}))$ 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 $\sum_{i=1}^n \xi_i^k (\Phi(x_i) - \Phi(x_{i-1}))$). 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 $\int_a^b f(u) d\Phi(u)$, como el límite de las sumas $\sum_{i=1}^n f(\xi_i) (\Phi(x_i) - \Phi(x_{i-1}))$. De esta forma surgió lo que ahora se conoce como la integral de Riemann-Stieltjes.

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 $f : [a, b] \rightarrow \mathbb{R}$, es integrable, entonces la función $x \rightarrow \int_a^x f(y) dy$, definida sobre $[a, b]$, es de variación acotada.

Definición 7. Se dice que una función $g : [a, b] \rightarrow \mathbb{R}$ es de **variación acotada** si:

$$\sup_P \sum_{k=1}^n |g(t_k) - g(t_{k-1})| < \infty$$

donde $P = \{a = t_0 < t_1 < \dots < t_n = b\}$ es una partición del intervalo $[a, b]$.

La relación entre la integral de Stieltjes y las funciones de variación acotada proviene de que una función $g : [a, b] \rightarrow \mathbb{R}$ 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.

Una propiedad muy importante de la integral de Stieltjes es la **fórmula de integración por partes**:

Supongamos que la función acotada $f : [a, b] \rightarrow \mathbb{R}$ es integrable con respecto a la función acotada $g : [a, b] \rightarrow \mathbb{R}$, entonces g es integrable con respecto a f y se tiene:

$$\int_a^b gdf = f(b)g(b) - f(a)g(a) - \int_a^b f dg$$

Por otra parte, para que una teoría de integración tenga alguna utilidad práctica, tal vez la mínima condición que se puede pedir es que todas las funciones continuas sean integrables y es aquí donde cobra su mayor importancia el concepto de función de variación acotada, el cual Lebesgue ya lo había identificado como muy importante. En ese sentido se tienen los siguientes resultados:

Sea $g : [a, b] \rightarrow \mathbb{R}$ una función de variación acotada y $f : [a, b] \rightarrow \mathbb{R}$ una función continua, entonces f es integrable con respecto a g .

Sea $g : [a, b] \rightarrow \mathbb{R}$ una función y supongamos que todas las funciones continuas $f : [a, b] \rightarrow \mathbb{R}$ son integrables con respecto a g , entonces g es de variación acotada.

Este último resultado es fundamental y se generalizaría, llegando a ser un concepto clave en la teoría general de la medida y la teoría general de integración. Podría decirse de la siguiente manera: **para tener una "buena" teoría de integración de una función f con respecto a otra función g , se requiere que g sea de variación acotada.**

Después de que Lebesgue desarrolló su teoría de integración en \mathbb{R} , se extendió al caso de \mathbb{R}^n sin mucha dificultad.

En 1913, Johann Karl August Radon 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.

Radon introdujo el concepto de funcional aditiva sobre una familia de subconjuntos de \mathbb{R}^n . Aún no se trataba de la definición general de una medida, la cual se dio más tarde.

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'intégrale d'une fonctionnelle étendue a un ensemble abstrait*.

Finalmente, Fréchet definió la integral con respecto a una función aditiva utilizando el método de Darboux, el cual consiste en definir la integral superior y la integral inferior.

En un 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.

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.

Nikodym retomó el trabajo de Radon y obtuvo un resultado general, ahora conocido como 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, \dots son elementos de la familia, ajenos por parejas, entonces $\mu(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$; es decir, μ es σ -aditiva, en la terminología moderna. En otras palabras, Nikodym trabajaba con medidas tal y como se les define actualmente.

Con el trabajo de Nikodym quedó formulada la teoría de la medida y quedaron establecidos los resultados básicos de la teoría de integración con respecto a una medida.

Si \mathbb{F} es un conjunto cualquiera y \mathfrak{S} es una σ -álgebra de subconjuntos de \mathbb{F} , una función $\mu : \mathfrak{S} \mapsto \overline{\mathbb{R}}$ es una medida si satisface las siguientes 3 propiedades:

1. $\mu(A) \geq 0$ para cualquier $A \in \mathfrak{S}$.
2. $\mu(\emptyset) = 0$
3. Si A_1, A_2, \dots es una familia infinita numerable de elementos de \mathfrak{S} , ajenos por parejas, entonces $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

Si éste es el caso, decimos que la terna $(\mathbb{F}, \mathfrak{S}, \mu)$ es un espacio de medida.

$\overline{\mathbb{R}}$ es el conjunto de los números reales extendido; es decir, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

La propiedad 3 es llamada la σ -aditividad de la medida; o bien se dice que μ es σ -aditiva.

3. TEORÍA DE LA PROBABILIDAD

Por el lado del Cálculo de Probabilidades, una de las propiedades básicas con la que se contaba para calcular probabilidades es la que formuló 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 $p_1 + p_2 + \dots + p_n$, donde, para $k \in \{1, 2, \dots, n\}$, p_k es la probabilidad de que se produzca de la k -ésima manera. ¡Es la aditividad finita!

¿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 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!

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. Hacia 1930 se tenía ya desarrollada una Teoría General de la medida, la cual incluía la definición y el estudio de medidas definidas sobre una familia de subconjuntos de un conjunto cualquiera y se contaba con un método para construir esas medidas, siendo la σ -aditividad la propiedad básica de cualquier medida.

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 seleccionada 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.

Se fueron planteando problemas de probabilidad cada vez más complejos, en particular 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 $f : [0, \infty) \rightarrow \mathbb{R}^2$; un espacio de dimensión infinita. Fue Norbert Wiener quien, en el año 1923, construyó ese modelo utilizando la moderna teoría de la medida y asumiendo que la función de probabilidad es σ -aditiva. Teníamos nuevamente la σ -aditividad de la función de probabilidad, pero todavía para un caso particular, aunque esta vez de mucha importancia.

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 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.

Lo que hizo Kolmogorov fue mostrar que el tratar a la función de probabilidad como una medida es consistente en el sentido de que, bajo condiciones mínimas, siempre es posible

construir una medida asociada a un fenómeno aleatorio. Así que, éste se modela con un espacio de medida $(\Omega, \mathfrak{S}, P)$, con la particularidad de que $P(\Omega) = 1$. En este caso, la terna $(\Omega, \mathfrak{S}, P)$ es llamada un espacio de probabilidad. Como siempre es posible extender la medida P de tal manera que todos los subconjuntos de los conjuntos de medida cero sean medibles, se la toma con esta característica, en cuyo caso se dice que el espacio de probabilidad es completo, o bien que la medida P es completa.

En este modelo, una variable aleatoria real es lo que en teoría de la medida se le conoce como función medible con valores en $\overline{\mathbb{R}}$, es decir, una función $X : \mathfrak{S} \mapsto \overline{\mathbb{R}}$ tal que $X^{-1}(B) \in \mathfrak{S}$ para cualquier subconjunto B de $\overline{\mathbb{R}}$ que pertenezca a la σ -álgebra generada por los intervalos en $\overline{\mathbb{R}}$, es decir, los conjuntos borelianos en $\overline{\mathbb{R}}$.

4. Procesos estocásticos

El estudio de los procesos estocásticos se inició a principios del siglo XX, en forma paralela a la conclusión del estudio de las sucesiones de variables aleatorias independientes. Sus principales precursores fueron A. A. Markov, A. N. Kolmogorov, A. Ya. Khintchine, N. Wiener y P. Lévy.

Desde la publicación del teorema de Bernoulli, en 1713, el motor de desarrollo de la teoría de la probabilidad fue la búsqueda de resultados que permitieran mejorar y generalizar ese teorema. Ese proceso culminó en la década de los 30's del siglo XX con la formulación acabada de los teoremas límite para sucesiones de variables aleatorias independientes.

Ley débil de los grandes números:

Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ . Entonces, para cualquier $\varepsilon > 0$, se tiene:

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{X_1 + \dots + X_n}{n} - \mu \right| > \varepsilon \right] = 0$$

Este tipo de convergencia es llamada **convergencia en probabilidad** y se suele escribir como sigue:

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} \mu$$

Ley fuerte de los grandes números:

Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ . Entonces:

$$P \left[\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mu \right] = 1$$

Este tipo de convergencia es llamada **convergencia casi segura o con probabilidad 1** y se suele escribir como sigue:

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{c.s.} \mu$$

Teorema del límite central:

Si X_1, X_2, \dots es una sucesión de variables aleatorias independientes e idénticamente distribuidas, de esperanza común μ y varianza común finita σ^2 , entonces, para cualquier $x \in \mathbb{R}$, se tiene:

$$\lim_{n \rightarrow \infty} P \left[\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy$$

Este tipo de convergencia es llamada **convergencia en distribución**.

Además de los tipos anteriores de convergencia de variables aleatorias, hay otro que es importante para el Cálculo Estocástico: la **convergencia en L^2** .

El conjunto de variables aleatorias X tales que $E[X^2] < \infty$, con la norma definida por $\|X\| = (E[X^2])^{\frac{1}{2}}$, forma un espacio de Hilbert, el cual es denotado por L^2 .

Se dice que una sucesión de variables aleatorias $(X_n)_{n \in \mathbb{N}}$ converge en L^2 a la variable aleatoria X si ésta y todas las variables aleatorias de la sucesión pertenecen a L^2 y la convergencia se da con la norma definida en L^2 ; es decir:

$$\lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0.$$

Si éste es el caso, se escribe $X_n \xrightarrow{L^2} X$.

Dentro de este contexto del estudio de los teoremas límite, Markov introdujo en 1906 una nueva línea de investigación al demostrar la ley débil de los grandes números para una sucesión de variables aleatorias no independientes, sino, de acuerdo con el término que él utilizó, encadenadas. Surgió así un tipo de proceso estocástico que ahora es llamado de Markov. De manera general e informal, se dice que un proceso estocástico es de Markov si dado cualquier tiempo t , las familias de variables aleatorias $\{X_s : s < t\}$ y $\{X_s : s > t\}$ son independientes, propiedad que en ocasiones se enuncia diciendo que, dado el presente, el pasado y el futuro son independientes. Por esta propiedad, también se dice que un proceso de este tipo es un proceso sin memoria.

Un gran avance en el estudio de los procesos estocásticos lo hizo Norbert Wiener entre 1921 y 1923 al construir un modelo matemático del movimiento browniano. A este modelo lo llamaremos un proceso de Wiener. Como veremos, la base del Cálculo Estocástico es precisamente este proceso.

Mientras tanto, la escuela rusa, que surgió con los trabajos de P. L. Chebyshev, realizó avances significativos generalizando el trabajo de Markov. Entre 1931 y 1937, Kolmogorov realizó un estudio a fondo de las cadenas de Markov en tiempo discreto y en tiempo continuo; también estudio los procesos de difusión, los cuales son procesos de Markov con espacios de

estados continuos. Por su parte, en el año 1934, Khintchine introdujo un tipo de procesos que pueden ser no markovianos: los procesos estacionarios.

Otro gran avance en el desarrollo de la teoría de los procesos estocásticos, lo realizó Paul Lévy en su libro *Théorie de l'addition des variables aléatoires*, cuya primera edición se publicó en 1937. Ahí planteó que una extensión natural de las sumas o series de variables aleatorias independientes es el siguiente: .En lugar de un índice entero n , introduciremos un parámetro t que varía de una manera continua; las sumas sucesivas S_n (se refiere a $S_n = \sum_{k=1}^n X_k$, donde X_1, X_2, \dots, X_n son variables aleatorias independientes) serán entonces reemplazadas por una función aleatoria $X(t)$ de ese parámetro, y la condición de independencia de los diferentes elementos cuya suma es $X(t)$ debe expresarse como sigue: si $t_0 < t_1$, el crecimiento $X(t_1) - X(t_0)$ es una variable aleatoria independiente de $X(t_0)$ y del conjunto de valores de $X(t)$ para $t \leq t_0$."(p. 158). Tomando $t \in \mathbb{R}^+$, lo anterior implica que si t_1, t_2, \dots, t_n son números reales no negativos tales que $t_1 < t_2 < \dots < t_n$, entonces las variables aleatorias $X(t_2) - X(t_1), X(t_3) - X(t_2), \dots, X(t_n) - X(t_{n-1})$ son independientes.

Lévy estaba definiendo ahí lo que se conoce ahora como un proceso con incrementos independientes. Aclaraba que algunos casos particulares de ese tipo de funciones aleatorias habían sido consideradas previamente por Bachelier, en 1913, Wiener, en 1923 y Kolmogorov, en 1932.

Obsérvese que si $(Y_n)_{n \in \mathbb{Z}^+}$ es un proceso con incrementos independientes, entonces, definiendo, para $k \in \mathbb{N}$, $X_n = Y_n - Y_{n-1}$, $(X_n)_{n \in \mathbb{N}}$ es una sucesión de variables aleatorias independientes y, para cualquier $n \in \mathbb{N}$, se tiene: $Y_n - Y_0 = \sum_{k=1}^n X_k$. De aquí que, para estudiar el proceso $(Y_n)_{n \in \mathbb{Z}^+}$, se puedan utilizar los resultados que se tienen para las sumas de variables aleatorias independientes.

En general, un proceso estocástico representa la dinámica de un fenómeno aleatorio; es decir un proceso estocástico es una familia de variables aleatorias $(X_t)_{t \in \Gamma}$, las cuales, para los fines de lo que veremos más adelante, las consideraremos con valores reales; de igual forma Γ es un conjunto de índices contenido en el conjunto de números reales. Las variables aleatorias X_t están definidas sobre un determinado espacio de probabilidad $(\Omega, \mathfrak{F}, P)$. El caso más común consiste en el estudio de un fenómeno aleatorio que evoluciona a partir de un tiempo inicial $t_0 \geq 0$, de manera que Γ es un subconjunto de $[0, \infty)$; en muchos casos de interés, se trata del conjunto de números enteros no negativos o del conjunto de números reales no negativos. En algunos casos es de interés considerar también una variable aleatoria X_∞ . Para cada $\omega \in \Omega$, la función $t \rightarrow X_t(\omega)$, definidas sobre Γ es llamada una **trayectoria** del proceso.

Denotaremos por \mathbb{Z}^+ al conjunto de números enteros no negativos y por \mathbb{R}^+ al conjunto de números reales no negativos. $\bar{\mathbb{Z}}^+$ denotará al conjunto $\mathbb{Z}^+ \cup \{\infty\}$ y $\bar{\mathbb{R}}^+$ al conjunto $\mathbb{R}^+ \cup \{\infty\}$. $\mathcal{B}(\mathbb{R})$ denotará la σ -álgebra de los conjuntos borelianos en \mathbb{R} .

Si $\Gamma = \mathbb{Z}^+$ diremos que $(X_t)_{t \in \Gamma}$ es un **proceso estocástico en tiempo discreto**, mientras que si $\Gamma = \mathbb{R}^+$ diremos $(X_t)_{t \in \Gamma}$ es un **proceso estocástico en tiempo continuo**.

Diremos que un proceso $(X_t)_{t \in \mathbb{R}^+}$ es continuo si existe un conjunto $C \in \mathfrak{S}$ de probabilidad 1, tal que, para cualquier $\omega \in C$, la función $t \rightarrow X_t(\omega)$, definida sobre \mathbb{R}^+ , es continua.

Para cada $t \in \Gamma$, X_t representa lo que llamaremos el estado del proceso en el tiempo t . Este estado no está determinado, puede ser uno cualquiera de los valores que toma X_t . Lo único conocido es la ley probabilística mediante la cual evoluciona el proceso, la cual, en particular, nos dice cuál es la distribución de X_t . Esta distribución no es más que una medida de probabilidad definida sobre la σ -álgebra generada por X_t , cuyos elementos son aquellos eventos que dependen únicamente del estado del proceso en el tiempo t , a saber, todos los de la forma $[X_t \in B]$, donde $B \in \mathcal{B}(\mathbb{R})$.

De manera más general, resulta de interés considerar, para cada $t \in \Gamma$, la σ -álgebra formada por todos aquellos eventos que dependen únicamente del estado del proceso, desde su inicio, hasta el tiempo t , es decir, la σ -álgebra generada por la familia de variables aleatorias $\{X_s : s \in \Gamma \text{ y } s \leq t\}$, la cual denotaremos por \mathfrak{S}_t . Esta familia es no decreciente en el sentido de que si $t_1, t_2 \in \Gamma$ y $t_1 < t_2$ entonces $\sigma\{X_s : s \in \Gamma \text{ y } s \leq t_1\} \subset \sigma\{X_s : s \in \Gamma \text{ y } s \leq t_2\}$. Una familia de este tipo será llamada una **filtración**.

Diremos que una filtración $\{\mathfrak{S}_t : t \in \Gamma\}$ es **completa** si, para cualquier $t \in \Gamma$, todos los eventos de probabilidad cero pertenecen a \mathfrak{S}_t .

Diremos que un proceso estocástico $(X_t)_{t \in \Gamma}$ está **adaptado a la filtración** $\{\mathfrak{S}_t : t \in \Gamma\}$ si, para cualquier $t \in \Gamma$, la variable aleatoria X_t es \mathfrak{S}_t -medible.

Si $(X_t)_{t \in \Gamma}$ es un proceso estocástico y $\{t_1, t_2, \dots, t_n\}$ es un subconjunto finito de Γ , la distribución del vector aleatorio $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ será llamada una **distribución finito dimensional** del proceso.

Recordemos que la distribución de un vector aleatorio (X_1, X_2, \dots, X_n) , formado por variables aleatorias reales, es el conjunto de probabilidades de la forma $P[(X_1, X_2, \dots, X_n) \in B]$, donde B es un conjunto boreliano de \mathbb{R}^n . Estas probabilidades quedan determinadas por las probabilidades de la forma $P[X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n]$, donde B_1, B_2, \dots, B_n son conjuntos borelianos de \mathbb{R} .

El siguiente resultado fue demostrado por Lévy y resultó ser de gran importancia.

Si $(X_t)_{t \in \mathbb{R}^+}$ es un proceso con incrementos independientes, entonces $(X_t)_{t \in \mathbb{R}^+}$ es continuo si y sólo si, para cualesquiera $s, t \in \mathbb{R}^+$ tales que $s < t$, la variable aleatoria $X_t - X_s$ tiene distribución normal.

Los ejemplos básicos, de procesos con incrementos independientes, que menciona Lévy en su libro de 1937, son el proceso de Wiener, el cual es continuo, y el proceso de Poisson, el cual es un proceso de saltos, sin discontinuidades fijas.

Con base en su resultado, Lévy mostró que las propiedades que caracterizan al proceso de Wiener, $(W_t)_{t \in \mathbb{R}^+}$, son las siguientes:

- i) $W_0 = 0$.
- ii) Si $0 < t_1 < \dots < t_n$, entonces las variables aleatorias $W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ son independientes.
- iii) Si $0 \leq s < t$, entonces la variable aleatoria $W_t - W_s$ tiene distribución normal de parámetros $\mu = 0$ y $\sigma^2 = t - s$.
- iv) $(W_t)_{t \in \mathbb{R}^+}$ es un proceso continuo.

Los resultados de la teoría de los procesos estocásticos obtenidos durante la primera mitad del siglo XX fueron recogidos, sistematizados y complementados con nuevos resultados, en dos libros que marcaron la pauta para continuar desarrollando esta teoría por lo menos durante los 30 años que les siguieron: El libro de Paul Lévy, *Processus stochastiques et mouvement brownien*, publicado en 1948 y el libro de Joseph Leo Doob, *Stochastic processes*, publicado en 1953.

Lévy comienza su libro dando una nueva construcción del proceso de Wiener y pasa después a considerar el problema general de la construcción de procesos estocásticos. En los capítulos siguientes expone la teoría de los procesos de Markov y de los procesos estacionarios. Después expone una versión más completa de su teoría de los procesos con incrementos independientes, llamados ahí procesos aditivos. Tal vez la parte más importante de su libro es el estudio a fondo que hace del proceso de Wiener, tanto en la recta real, como en el plano y en otros espacios. Ahí mostró que las trayectorias de un proceso de Wiener son, con probabilidad 1, funciones no diferenciables y, por lo tanto, no son de variación acotada.

El libro de J. L. Doob fue fuente de inspiración para muchos matemáticos ya que ahí plantea varias líneas de investigación para diferentes problemas. Doob sigue lo que podríamos llamar el enfoque moderno: se parte de los axiomas de Kolmogorov y la construcción de procesos se realiza vía el teorema de extensión de Kolmogorov. Comienza su libro estudiando propiedades generales de los procesos estocásticos, en particular, la existencia de versiones de un proceso con trayectorias que son regulares en algún sentido. Realiza un estudio a fondo de los procesos de Markov en tiempo discreto y en tiempo continuo. Trata también el estudio de los procesos con incrementos independientes y de los procesos estacionarios. Tal vez las principales aportaciones de Doob en ese libro son, por una parte, el estudio que hace de los procesos conocidos como martingalas, los cuales adquirirían una importancia central dentro de la teoría de los procesos estocásticos, y, por otra parte, la invención del concepto de tiempo de paro, el cual es un tiempo aleatorio. Paul Meyer, en la segunda edición de su libro dice al respecto (p. 184): "Il a sans doute fallu autant de génie aux créateurs du calcul infinitésimal pour expliciter la notion si simple de dérivée, qu'à leurs successeurs pour faire toute le reste. L'invention des temps d'arrêt par Doob est tout a fait comparable" (Sin duda, los creadores del cálculo infinitesimal requirieron tanto genio para explicar la simple noción de derivada como a sus sucesores hacer el resto. La invención de los tiempos de paro por Doob es totalmente comparable).

La teoría que más se desarrolló en los años posteriores a la publicación de los libros de Lévy y de Doob fue la de los procesos de Markov, en particular, la de los procesos de difusión. Uno de los problemas centrales que se atacó fue el de la construcción de procesos con trayectorias regulares en algún sentido.

La escuela rusa, con Kolmogorov a la cabeza, hizo las más importantes contribuciones, desarrollando ampliamente la teoría de los procesos de Markov. Uno de los más destacados en este periodo fue E. B. Dynkin, quien vinculó la teoría de los semigrupos de operadores con la de los procesos de Markov.

No se puede dejar de mencionar a G. A. Hunt por sus trabajos en la teoría de los procesos de Markov y su relación con la teoría del potencial y a K. L. Chung por sus aportaciones a la teoría de los procesos de Markov y por su incansable labor de difusión en varios libros de la teoría de la probabilidad y de los procesos estocásticos.

En los 60's comenzó a desarrollarse la escuela francesa, con Paul Meyer a la cabeza. En el año 1966 se publicó el primero de sus libros, titulado *Probabilités et Potentiel*, en el cual inició el desarrollo de lo que se llamaría después la teoría general de los procesos estocásticos. En 1967 se publicó otro libro suyo, titulado *Processus de Markov*, en el cual presenta algunos de los capítulos que incluiría la segunda parte de su primer libro. Sin embargo, el rumbo de las investigaciones cambió y no hubo tal segunda parte; en su lugar, en la segunda edición de su primer libro, publicada en 1975 y teniendo como coautor a Claude Dellacherie, hubo una amplia reformulación, a tal grado que más bien esa segunda edición puede verse como un libro distinto. De esa segunda edición sí hubo una segunda parte, titulada *Théorie des martingales*.

Son interesantes los comentarios de Meyer y Dellacherie en el prefacio de la segunda edición de *Probabilités et Potentiel*. Para empezar, hacen la aclaración de que, a pesar del título, el libro contiene poco de probabilidad y nada de teoría del potencial. La razón para ir en otra dirección fue, como ellos lo dicen, la evolución tan rápida de la teoría, con lo cual se crearon nuevos conceptos y algunas cosas que parecían muy importantes antes, ya no lo eran tanto, mientras que aspectos que parecían sin interés, se volvieron importantes. Como enseñanza, dicen los autores: "Hemos tratado en particular de liberarnos de la actitud consistente en creer que un texto matemático presenta, para nuestra contemplación, verdades eternas, salidas del mundo de las ideas, y cuyo valor no puede sufrir alguna inflación... Pero las verdades existen no por el hecho que se impriman sobre buen papel, sino por el interés que se les da. Muchas verdades inmutables de 1966, están muertas, mientras que pequeños comentarios de 1966 aclaran ahora grandes partes de la teoría."

Esto es un ejemplo de cómo la matemática se va creando como producto social y no mediante "descubrimientos" de "verdades" que se encuentran escondidas en alguna parte, las cuales hay que ir encontrando.

5. ECUACIONES INTEGRALES ESTOCÁSTICAS

En su libro *Processus stochastiques et mouvement brownien* (1948), Lévy trató el problema de la construcción de procesos mediante la solución de ecuaciones diferenciales estocásticas. La idea que formuló es que el incremento infinitesimal de un proceso $(X_t)_{t \in \mathbb{R}^+}$ en un tiempo t , lo cual denota por $dX(t)$, podría estar dado por una función que depende de t , dt , $X(t)$ y una o varias variables aleatorias de tipo familiar, como, por ejemplo, variables aleatorias independientes con distribución normal. En el caso del proceso de Wiener $(W_t)_{t \in \mathbb{R}^+}$, como el incremento $W_t - W_s$, para $s, t \in \mathbb{R}^+$ tales que $s < t$, tiene distribución normal con esperanza cero y varianza $t - s$, entonces en un incremento dt del tiempo, el incremento $dW(t)$ tiene distribución normal con esperanza cero y varianza dt , de manera que, denotando por ξ a una variable aleatoria con distribución normal estándar, el proceso de Wiener puede definirse por la ecuación diferencial estocástica $dW(t) = \xi \sqrt{dt}$.

Analizó algunos casos particulares de ecuaciones diferenciales estocásticas y planteó el problema general de resolver una ecuación diferencial estocástica del tipo siguiente:

$$dX(t) = f(t, X, Y, Z) dY(t) + g(t, X, Y, Z) dZ(t)$$

donde $(Y_t)_{t \in \mathbb{R}^+}$ y $(Z_t)_{t \in \mathbb{R}^+}$ son procesos estocásticos definidos previamente.

Hacía énfasis en que hay una dificultad para resolver una ecuación de este tipo cuando $Y(t)$ y $Z(t)$ no son funciones de variación acotada, por ejemplo cuando Y y Z representan las coordenadas de un movimiento browniano en el plano. Mostró entonces que una integral del tipo $\int_{t_0}^{\tau} f(t, X, Y, Z) dY(t)$, a la cual denominaba integral estocástica, puede evaluarse no siguiendo el método clásico de la integral de Riemann, sino considerando números τ_1, τ_2, \dots , elegidos al azar en el intervalo (t_0, τ) y tomando después, para cada $n \in \mathbb{N}$, los primeros n de esos números, ordenados del menor al mayor, para formar con ellos una suma de tipo Riemann. Consideró el caso del cálculo de áreas estocásticas, es decir, áreas de curvas definidas por un proceso estocástico y mostró que las integrales estocásticas que resultan quedan bien definidas con el método mencionado.

Siguiendo un trabajo de S. Bernstein, consideró los procesos de Markov que son solución a ecuaciones diferenciales estocásticas de la forma:

$$dX(t) = A[t, X_t] dt + B[t, X_t] \xi \sqrt{dt}$$

donde ξ es una variable aleatoria auxiliar que cumple con determinadas condiciones.

Obsérvese que en el caso en que ξ es una variable aleatoria con distribución normal estándar, la ecuación anterior puede escribirse en la siguiente forma:

$$dX(t) = A[t, X_t] dt + B[t, X_t] dW_t$$

Planteó entonces el problema general de la determinación de todos los procesos de Markov como soluciones de ecuaciones diferenciales estocásticas. Señalaba que ese problema no estaba

resuelto de manera general, pero dio la solución para el caso de los procesos con incrementos independientes.

La dificultad central para tratar con una integral del tipo $\int_0^t B[s, X_s] dW_s$ consiste en que las funciones $t \rightarrow W_t$ no son de variación acotada y entonces no pueden definirse como integrales de Lebesgue.

Además, se tiene el siguiente resultado:

$$\lim_{\|P\| \rightarrow 0} \sum_{k=1}^n (W_{t_k} - W_{t_{k-1}})^2 = b - a \text{ en } L^2$$

donde $P = \{a = t_0 < t_1 < \dots < t_n = b\}$ es una partición del intervalo $[a, b]$.

Al límite $\lim_{\|P\| \rightarrow 0} \sum_{k=1}^n (W_{t_k} - W_{t_{k-1}})^2$ se le llama la **variación cuadrática de (W_t)** en el intervalo $[a, b]$.

Utilizando este resultado, podemos ver uno de los problemas que existen para definir una integral del tipo $\int_0^t B[s, X_s] dW(t)$ de la manera usual, como integral de Riemann. Por ejemplo, si quisiéramos definir la integral $\int_0^t W_s dW_s$ nos enfrentamos al siguiente problema:

$$\lim_{\|P\| \rightarrow 0} \sum_{k=1}^n W_{t_{k-1}} [W_{t_k} - W_{t_{k-1}}] = \frac{1}{2}W_t^2 - \frac{1}{2}t \text{ en } L^2$$

$$\lim_{\|P\| \rightarrow 0} \sum_{k=1}^n W_{t_k} [W_{t_k} - W_{t_{k-1}}] = \frac{1}{2}W_t^2 + \frac{1}{2}t \text{ en } L^2$$

donde $P = \{0 = t_0 < t_1 < \dots < t_n = t\}$ es una partición del intervalo $[0, t]$.

6. MARTINGALAS

Un concepto central en el Cálculo Estocástico es el de Martingala, el cual, para su misma definición requiere del concepto de esperanza condicional dada una σ -álgebra. Éste fue introducido por Kolmogorov en su libro de 1933 utilizando uno de los resultados más impactantes de la teoría de la medida, el teorema de Radon-Nikodym. Se fue gestando desde los trabajos de Lebesgue y Radon, de 1904 y 1913, respectivamente, pero fue Nikodym quien lo demostró con toda claridad y generalidad en su artículo de 1930.

Definición 8. Sean μ y ν dos medidas definidas sobre el mismo espacio $(\mathbb{F}, \mathfrak{S})$, donde \mathfrak{S} es una σ -álgebra de subconjuntos de \mathbb{F} . 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 $\mu(E) = 0$.

Teorema de Radon-Nikodym:

Sean μ y ν dos medidas definidas sobre el mismo espacio $(\mathbb{F}, \mathfrak{S})$. 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 .

Utilizando este resultado se demuestra el siguiente, el cual permite definir la esperanza condicional de una variable aleatoria dada una σ -álgebra

Sea X una variable aleatoria de esperanza finita definida sobre un espacio de probabilidad $(\Omega, \mathfrak{S}, P)$ y \mathcal{G} una sub σ -álgebra de \mathfrak{S} , existe entonces una variable aleatoria Z , medible con respecto a \mathcal{G} , de esperanza finita y tal que $\int_B Z dP = \int_B X dP$ para cualquier conjunto $B \in \mathcal{G}$. Además, si Y es otra variable aleatoria con las mismas propiedades que Z , entonces $P[Y = Z] = 1$.

Definición 9. Sea X es una variable aleatoria de esperanza finita y \mathcal{G} una sub σ -álgebra de \mathcal{F} . Se dice que Z es una versión de la **esperanza condicional de X con respecto a \mathcal{G}** y se escribe $E[X | \mathcal{G}] = Z$ c.s. si Z una variable aleatoria medible con respecto a \mathcal{G} , de esperanza finita y tal que $\int_B Z dP = \int_B X dP$ para cualquier evento $B \in \mathcal{G}$.

Definición 10. Dado un espacio de probabilidad completo $(\Omega, \mathfrak{S}, P)$, un conjunto de índices $\Gamma \subset \mathbb{R}$ y una filtración $\{\mathfrak{S}_t : t \in \Gamma\}$, se dice que un proceso $(M_t)_{t \in \Gamma}$ es una **martingala** con respecto a la filtración $\{\mathfrak{S}_t : t \in \Gamma\}$ si se satisfacen las siguientes tres propiedades:

1. El proceso $(M_t)_{t \in \Gamma}$ está adaptado a la filtración $\{\mathfrak{S}_t : t \in \Gamma\}$.
2. $E[|M_t|] < \infty$ para cualquier $t \in \Gamma$.
3. Dados $s, t \in \Gamma$, con $s < t$, se tiene $E[M_t | \mathfrak{S}_s] = M_s$.

7. LA INTEGRAL DE ITO

Si $(X_t)_{t \in \mathbb{R}^+}$ es un proceso estocástico, denotaremos por X a la función $X : \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ definida por $X(t, \omega) = X_t(\omega)$. Inversamente, si $X : \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ es una función tal que, para cada $t \in \mathbb{R}^+$ la función $\omega \rightarrow X(t, \omega)$ es una variable aleatoria, denotaremos por $(X_t)_{t \in \mathbb{R}^+}$ al proceso estocástico formado por las variables aleatorias $X_t : \Omega \rightarrow \mathbb{R}$, definidas por $X_t(\omega) = X(t, \omega)$. Si C es un subconjunto de $\mathbb{R}^+ \times \Omega$ y $t \in \mathbb{R}^+$, denotaremos por C_t a la función $\omega \rightarrow I_C(t, \omega)$, definida sobre Ω .

Una ecuación integral estocástica es una relación de la forma:

$$X_t = X_0 + \int_0^t \mu(X_t, t) dt + \int_0^t \sigma(X_t, t) dW_t$$

donde $(X_t)_{t \geq 0}$ es un proceso estocástico, μ y σ son dos funciones reales definidas en \mathbb{R}^2 y $(W_t)_{t \geq 0}$ es un proceso de Wiener. Ecuaciones de este tipo se presentan en una diversidad de situaciones. La idea general es que si un sistema dinámico es modelado por una ecuación diferencial ordinaria $\frac{dX}{dt} = \mu(t, X)$ y dicho sistema es perturbado por la presencia de un

ruido. aleatorio, entonces la modelación estaría dada por una ecuación de la forma $\frac{dX}{dt} = \mu(t, X) + \sigma(t, X)\xi(t)$, en donde ξ representa la perturbación aleatoria. El ruido ξ se modela usualmente como la "derivada" de un proceso de Wiener, de manera que la última ecuación se escribiría en la forma:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t$$

μ es llamada el **arrastre** (drift) y σ la **volatilidad**.

Se sabe que con probabilidad 1, **las funciones $t \rightarrow W_t$ no son diferenciables**, así que la ecuación anterior es en realidad una ecuación integral:

$$X_t = X_0 + \int_0^t \mu(X_s, s)ds + \int_0^t \sigma(X_s, s)dW_s$$

También sabemos que, con probabilidad 1, las funciones $t \rightarrow W_t$ no son de variación acotada en ningún intervalo finito. No es posible entonces definir la integral estocástica $\int_0^t \sigma(X_s, s)dW_s$ como una integral de Lebesgue.

A pesar de las dificultades planteadas para definir la integral con respecto al proceso de Wiener, K. Ito (Stochastic Integral, 1944) logró darle un sentido preciso a una integral de la forma $\int_0^t Z_s dW_s$ para una familia amplia de procesos $(Z_t)_{t \geq 0}$, entre los que se incluyen a todos los procesos adaptados y continuos.

La idea general para definir la integral estocástica con respecto al proceso de Wiener es la siguiente:

Lo primero es que el proceso de Wiener $(W_t)_{t \geq 0}$ tiene las siguientes dos propiedades:

1. El proceso $(W_t)_{t \geq 0}$ es una martingala.
2. El proceso $W_t^2 - t$ es una martingala.

El proceso que dio Ito para definir la integral estocástica con respecto al proceso de Wiener es interesante ya que lo inicia definiendo la integral estocástica para una determinada familia de procesos estocásticos utilizando la integral de Riemann-Stieltjes:

Dada cualquier función $U : \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ tal que, para cada $\omega \in \Omega$, la función $t \rightarrow U(t, \omega)$ sea de variación acotada sobre los intervalos compactos, la integral $\int_0^t U(s, \omega) dW_s(\omega)$ está bien definida, para cualquier $\omega \in \Omega$, como integral de Riemann-Stieltjes. Denotaremos a esa integral como $\int_0^t U_s dW_s$.

Esto ocurre en particular cuando, para cada $\omega \in \Omega$, la función $t \rightarrow U(t, \omega)$ es escalonada sobre cualquier intervalo de la forma $[0, u]$.

Así que, si $Y : \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ tiene la forma:

$$Y = b_0 I_{\{0\} \times F_0} + \sum_{k=1}^m b_k I_{(s_k, t_k] \times F_k} + b_{m+1} I_{(s_{m+1}, \infty) \times F_{m+1}},$$

donde $b_0, b_1, \dots, b_m, b_{m+1}$ son números reales, $F_0, F_1, \dots, F_m, F_{m+1}$ pertenecen a \mathfrak{F} y $s_1, t_1, s_2, t_2, \dots, s_m, t_m, s_{m+1}$ son números reales no negativos, la integral $\int_0^t U(s, \omega) dW_s(\omega)$ está bien definida para cualquier $t \in \mathbb{R}^+$ y cualquier $\omega \in \Omega$. Obsérvese que en cada una de estas integrales ω está fija; se integra sobre s .

Si $t \in (s_n, t_n]$, entonces:

$$\int_0^t U(s, \omega) dW_s(\omega) = \sum_{k=1}^{n-1} b_k (W_{t_k} - W_{s_k}) I_{F_k}(\omega) + b_n (W_t - W_{s_n}) I_{F_n}(\omega)$$

Cuando $F_0 \in \mathfrak{F}_0$ y, para cualquier $k \in \{1, \dots, m, m+1\}$, $F_k \in \mathfrak{F}_{s_k}$, vamos a decir que una función U de ese tipo es **elemental**. Al conjunto de funciones elementales lo denotaremos por \mathcal{E} .

A la σ -álgebra de subconjuntos de $\mathbb{R}^+ \times \Omega$ generada por las funciones elementales la vamos a denominar la **σ -álgebra previsible** y la vamos a denotar por \mathcal{P} .

Se demuestra que si U es una función elemental, entonces el proceso $(\int_0^t U_s dW_s)_{t \in \mathbb{R}^+}$ es una martingala continua y nula en $t = 0$. Esto viene del hecho de que $(W_t)_{t \geq 0}$ es una martingala continua y nula en 0.

Además, si definimos $N_t = \int_0^t U_u dW_u$, el proceso $(N_t^2 - \int_0^t U_s^2 ds)_{t \in \mathbb{R}^+}$ resulta ser una martingala continua y nula en $t = 0$.

Por otra parte, para cada $t \in \mathbb{R}^+$, la función $C \rightarrow E \left[\int_0^t C_s ds \right]$, definida sobre \mathcal{P} , es una medida finita la cual denotaremos por μ_t .

Se demuestra entonces que el conjunto de las funciones elementales es denso en el espacio $L^2(\mathbb{R}^+ \times \Omega, \mathcal{P}, \mu_t)$ y que la función $G_t : I_{[0,t]} U \rightarrow \int_0^t U_s dW_s$, definida sobre \mathcal{E} , es una isometría lineal, es decir, es una función lineal y se tiene:

$$\|I_{[0,t]} U\|_{L^2(\mathbb{R}^+ \times \Omega, \mathcal{P}, \mu_t)} = \left\| \int_0^t U_s dM_s \right\|_{L^2(\Omega, \mathfrak{F}, P)}$$

Como \mathcal{E} es denso en $L^2(\mathbb{R}^+ \times \Omega, \mathcal{P}, \mu_t)$, Si $(Z_t)_{t \in \mathbb{R}^+}$ es un proceso previsible en $L^2(\mathbb{R}^+ \times \Omega, \mathcal{P}, \mu_t)$, esta isometría lineal se extiende a una isometría lineal G_t definida sobre todo $L^2([0, t] \times \Omega, \mathcal{P}, \mu_t)$. Esta extensión asocia entonces a cada elemento $X \in L^2([0, t] \times \Omega, \mathcal{P}, \mu_t)$ una variable aleatoria $G_t(X) \in L^2(\Omega, \mathfrak{F}, P)$.

Así que para cada $t \in \mathbb{R}^+$, tenemos una variable aleatoria $M_t \in L^2(\Omega, \mathfrak{F}, P)$, sin embargo, el proceso $(M_t)_{t \in \mathbb{R}^+}$ no necesariamente es una martingala y no necesariamente es un proceso continuo; pero se tiene el siguiente resultado:

Si $(Z_t)_{t \in \mathbb{R}^+}$ es un proceso previsible en $L^2_{\mathcal{P}}(\mu_t)$, existe una única martingala continua nula en cero, $(N_t)_{t \in \mathbb{R}^+}$, tal que:

i) Si $(Z^{(n)})_{n \in \mathbb{N}}$ es una sucesión de procesos elementales que converge a Z en $L^2_{\mathcal{P}}(\mu_t)$, entonces, para cualquier $t \in \mathbb{R}^+$, la sucesión $(\int_0^t Z_s^{(n)} dW_s)_{n \in \mathbb{N}}$ converge a N_t en $L^2(\Omega, \mathfrak{F}, P)$.

ii) El proceso $(N_t^2 - \int_0^t Z_s^2 ds)_{t \in \mathbb{R}^+}$ es una martingala continua nula en cero.

El proceso $(\int_0^t Z_s^2 ds)_{t \in \mathbb{R}^+}$ es llamado el compensador de la martingala $(N_t)_{t \in \mathbb{R}^+}$ y se le denota por $(\langle N, N \rangle_t)_{t \in \mathbb{R}^+}$.

Una vez definida la integral estocastica $N_t = \int_0^t Z_s dW_s$ para un proceso previsible $(Z_t)_{t \in \mathbb{R}^+}$, de tal forma que el proceso $(N_t)_{t \in \mathbb{R}^+}$ sea una martingala continua, si $(Y_t)_{t \in \mathbb{R}^+}$ es otro proceso previsible, se puede definir $X_t = \int_0^t Y_s dN_s$ como $\int_0^t Y_s Z_s dW_s$ (con las condiciones que se requieran para que esta integral quede bien definida).

Finalmente, Ito demostró una relación que permite el manejo de las integrales estocásticas:

Si $F : \mathbb{R} \rightarrow \mathbb{R}$ es una función de clase C^2 , entonces:

$$F(X_t) = \int_0^t F'(X_s) dX_s + \frac{1}{2} \int_0^t F''(X_s) d\langle X, X \rangle_s$$

Esta relación es conocida como la **fórmula de Ito**.

8. EXTENSIÓN DE LA INTEGRAL DE ITO

En su libro *Stochastic processes*, Doob conjeturó que, si dada una martingala $(M_t)_{t \in \mathbb{R}^+}$ se pudiera encontrar un proceso no decreciente $(A_t)_{t \in \mathbb{R}^+}$ tal que el proceso $(M_t^2 - A_t)_{t \in \mathbb{R}^+}$ sea una martingala, podría definirse una integral estocástica con respecto a la martingala $(M_t)_{t \in \mathbb{R}^+}$ vía su proceso no decreciente $(A_t)_{t \in \mathbb{R}^+}$ asociado, al igual que se define la integral estocástica con respecto al proceso de Wiener $(W_t)_{t \in \mathbb{R}^+}$ gracias a que este proceso es una martingala y el proceso $(W_t^2 - t)_{t \in \mathbb{R}^+}$ también es martingala.

Paul André Meyer demostró que efectivamente, bajo determinadas condiciones, es posible realizar la conjetura de Doob y así definir la integral estocástica con respecto a una martingala. Una vez demostrado esto, la escuela francesa desarrolló una teoría general de la integral estocástica y más aún, una teoría general de los procesos estocásticos.

La condición que se requiere para asociar un proceso no decreciente a una martingala $(M_t)_{t \in \mathbb{R}^+}$ consiste en que ésta sea **acotada en $L^2(\Omega, \mathfrak{F}, P)$** , lo cual significa lo siguiente:

$$\sup \{E[M_t^2] : t \in \mathbb{R}^+\} < \infty$$

El resultado de Meyer, después de algunas correcciones que se hicieron posteriormente, quedó como sigue:

Si $(M_t)_{t \in \mathbb{R}^+}$ es una martingala acotada en $L^2(\Omega, \mathfrak{F}, P)$, entonces existe un único proceso $(A_t)_{t \in \mathbb{R}^+}$ adaptado, no decreciente y nulo en 0 con las siguientes propiedades:

i) $A_\infty(\omega) = \lim_{t \rightarrow \infty} A_t(\omega) < \infty$ para cualquier $\omega \in \Omega$.

ii) El proceso $(M_t^2 - A_t)_{t \in \mathbb{R}^+}$ es una martingala.

El proceso $(A_t)_{t \in \mathbb{R}^+}$ es llamado el **compensador** de la martingala $(M_t)_{t \in \mathbb{R}^+}$.

Es gracias a la existencia de este proceso $(A_t)_{t \in \mathbb{R}^+}$ que se puede definir la integral estocástica $\int_0^t Z_s dM_s$, ya que las integrales con respecto a A_t (como función de t) están bien definidas como integrales de Lebesgue.