Sobre la Axiomatización de la Probabilidad

Ezequiel Castro Valenzuela[1]
Escuela de Ciencias Físico Matemáticas
Universidad Autónoma de Sinaloa

Contenido

[1]Este trabajo se llevó a cabo bajo la asesoría del Dr. Víctor M. Pérez Abreu C. y es parte del Programa XVIII Verano de la Investigación Científica de la AMC y del Primer Verano de Probabilidad y Estadística del CIMAT, 2008.

1 Introducción

Al igual que otras ramas de la matemática como la geometría, el álgebra y la topología, la teoría moderna de la probabilidad está basada en axiomas. En este trabajo se mencionan algunos aspectos y circunstancias del proceso de axiomatización de la probabilidad. Se hace énfasis en el estado de esta disciplina a finales del Siglo XIX, así como los numerosos matemáticos que intervinieron en el proceso de axiomatización de la probabilidad.

En la Sección 2 se exponen algunos conceptos sobre probabilidad clásica tales como la definición de probabilidad clásica y un par de reglas (teoremas) importantes que se derivan de la definición. Se discute el principio de Cournot, el cual sugiere un puente entre la probabilidad matemática y el mundo empírico. También se presentan algunas paradojas.

La Sección 3 trata sobre el Sexto Problema de Hilbert y sobre algunas respuestas de varios matemáticos a este problema. También mencionamos el *Grundlagen der Geometrie*, obra de Hilbert en la cual muestra las características que debe cumplir un sistema axiomático.

En la Sección 4 se hace mención de algunos matemáticos cuyas contribuciones favorecieron el desarrollo de la teoría moderna de probabilidad. Veremos que muchas de estas contribuciones tienen que ver con la teoría matemática conocida como teoría de la medida. Al avanzar en esta sección veremos que, lejos de que la axiomatización de la probabilidad sea un logro de Kolmogorov sólo, ésta se dio sólo después de algunas contribuciones e ideas importantes dadas por otros matemáticos, entre ellos Borel, Lebesgue y Fréchet. Kolmogorov mismo reconoció este hecho; por ejemplo, en el segundo párrafo del prefacio de *Fundamentos* encontramos la frase: "Esta tarea hubiera sido algo imposible antes de la introducción de las teorías de Lebesgue de medida e integración." (Por el término *Fundamentos* nos referimos al texto de 1933 en el que Kolmogorov propone sus axiomas; el nombre completo en alemán es *Grundbegriffe der Wahrscheinlichkeitsrechnung*.)

Finalmente, en la Sección 5 fijamos nuestra atención en *Fundamentos*. Se mencionan los seis axiomas, enfoncándonos en el sexto axioma o axioma de continuidad. Veremos cómo este axioma se relaciona con la aditividad numerable.

2 Probabilidad Clásica

La definición de probabilidad adoptada por Jacob Bernoulli en *Ars Conjectandi* de 1713, y por Abraham de Moivre en *The Doctrine of Chances* de 1718 es:

La probabilidad de un evento es la razón del número de casos igualmente probables que lo favorecen entre el número total de casos posibles igualmente probables bajo las circunstancias.

2.1 Reglas de la Probabilidad Clásica

A partir de esta definición, de Moivre dedujo las siguientes dos reglas de probabilidad:

Teorema 1 *Teorema de probabilidad total o teorema de adición: si A y B no pueden ocurrir simultáneamente, entonces*

$$\textit{probabilidad de la ocurrencia de } A \textit{ o de } B = \frac{\#\textit{ de casos que favorecen } A \textit{ o } B}{\#\textit{ total de casos}}$$

$$= \frac{\#\textit{ de casos que favorecen } A}{\#\textit{ total de casos}} + \frac{\#\textit{ de casos que favorecen } B}{\#\textit{ total de casos}}$$

$$=(\textit{probabilidad de } A) + (\textit{probabilidad de } B).$$

.

Teorema 2 *Teorema de probabilidad compuesta o teorema de multiplicación:*

$$\textit{probabilidad de la ocurrencia de } A \textit{ y } B = \frac{\#\textit{ de casos que favorecen } A \textit{ y } B}{\#\textit{ total de casos}}$$

$$= \frac{\#\textit{ de casos que favorecen } A}{\#\textit{ total de casos}} \cdot \frac{\#\textit{ de casos que favorecen } A \textit{ y } B}{\#\textit{ de casos que favorecen } A}$$

$$= (\textit{probabilidad de } A) \cdot (\textit{probabilidad de } B \textit{ si } A \textit{ ocurre})$$

En los siglos XVIII y XIX, la teoría de la probabilidad tuvo importantes avances y dio a luz importantes resultados, los cuales se podían obtener a partir del concepto de probabilidad definido en términos de eventos igualmente probables.

Uno de estos resultados importantes es el conocido como "Ley de los Grandes Números". Jacob Bernoulli formuló y dio una demostración de este teorema en *Ars Conjectandi*. Este teorema nos dice que en una sucesión suficientemente grande de ensayos independientes de un evento, existe una probabilidad muy alta de que la frecuencia con la que sucede el evento estará cerca de su probabilidad. La formulación matemática moderna de esta ley (la ley débil de los grandes números) es como sigue: consideremos un evento cuya probabilidad de ocurrencia es p, y una sucesión de n ensayos, y denotemos por S_n el número de ocurrencias del evento en los n ensayos, entonces, para cada $\delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| < \delta\right) = 1. \tag{1}$$

2.2 La Probabilidad Geométrica

Desde los mismos inicios del desarrollo de la teoría de la probabilidad se hizo notorio que la definición de probabilidad clásica basada en la consideración de un grupo finito de eventos equiprobables era insuficiente. Incluso en aquel tiempo algunos ejemplos especiales llevaron a una cierta modificación de la definición y a la construcción de un concepto de probabilidad para los casos en los cuales es concebible un conjunto infinito de resultados. Tal como antes, el concepto de eventos equiprobables juega un papel básico.

El problema general que se plantea al respecto y que conduce a una extensión de la noción de probabilidad puede ser formulado de la siguiente manera.

Consideremos una determinada región G del plano y, dentro de ella, otra región g con una frontera rectificable. Se lanza un punto al azar sobre G y deseamos encontrar la probabilidad de que el punto caiga dentro de la región g. Aquí la expresión "se lanza un punto al azar sobre la región G" significa que: el punto lanzado puede caer en cualquier punto de G; la probabilidad de que el punto caiga en cualquier porción de la región G es proporcional al área de esa porción y es independiente de su posición y de su forma.

Así, por definición, la probabilidad de que un punto lanzado al azar sobre G caiga en g es igual a

$$p = \frac{\acute{a}rea(g)}{\acute{a}rea(G)}.$$

La noción de probabilidad geométrica se extiende para abarcar otras medidas como longitud y volumen. De hecho, fue a principios del siglo XX que Borel y Lebesgue extendieron la clase de conjuntos sobre los cuales podemos definir extensiones geométricas.

Podemos hallar más sobre la probabilidad geométrica en el libro de texto de Gnedenko.

2.3 El Principio de Cournot

"El evento físicamente imposible es aquél cuya probabilidad tiende a ser infinitamente pequeña." (Cournot, 1843).

Existen la forma fuerte y la forma débil del principio de Cournot. (Fréchet, 1951, página 6; Martin, 2003). La forma fuerte se refiere a un evento con probabilidad muy pequeña o probabilidad cero, elegido de antemano de un ensayo en particular: nos dice que el evento no sucederá en ese ensayo. La forma débil nos dice que un evento con probabilidad muy pequeña sucederá muy raramente en ensayos repetidos.

Borel, Lévy y Kolmogorov combinaron el principio de Cournot con el teorema de Bernoulli (de los grandes números) para producir la conclusión inequívoca de que la probabilidad de un evento *será* aproximada por su frecuencia en una sucesión particular de ensayos independientes suficientemente grande. Esto requiere la forma fuerte del principio de Cournot.

Otros autores como Chuprov y Fréchet aplicaron el principio de Cournot de manera distinta. Para ellos, el principio de Cournot combinado con el teorema de Bernoulli nos lleva a la conclusión de que la probabilidad de un evento *usualmente* será aproximada por su frecuencia en una sucesión suficientemente grande. Sin embargo, cuando estimamos una probabilidad a partir de una frecuencia observada estamos dando un paso más adelante: estamos suponiendo que lo que usualmente sucede ha sucedido en esta observación en particular. Este paso requiere la forma fuerte del principio de Cournot.

El principio de Cournot sugiere el puente entre la probabilidad matemática y el mundo de la experiencia al declarar como imposibles aquellos eventos cuya probabilidad tiende a ser infinitamente pequeña. Lévy hizo claro este principio en su *Calcul des Probabilités* (1925) al explicar que la noción de eventos equiprobables es suficiente para la parte matemática de la probabilidad, pero en tanto basemos nuestro razonamiento en esta sola noción, nuestras probabilidades son meramente subjetivas. La noción del evento muy poco probable es lo que permite que los resultados de la teoría matemática adquieran importancia práctica. Combinando la noción de un evento muy poco probable con el teorema de Bernoulli, obtenemos la noción de probabilidad objetiva de un evento, una constante física que se mide por la frecuencia.

Por otra parte, matemáticos ingleses, entre ellos John Venn (1888), reconocieron la utilidad de la probabilidad, pero insistieron en definirla directamente en términos de frecuencias, dejando al teorema de Bernoulli y al principio de Cournot sin desempeñar papel alguno.

2.4 Las Paradojas de Bertrand

¿Cómo determinar cuándo dos eventos son equiprobables? Esta es una pregunta natural que suscita cierta discusión en cualquier época. Las décadas previas a *Fundamentos* no fueron la excepción.

En lo que respecta a la teoría de la probabilidad, decidir cuándo dos eventos son equiprobables no era en realidad el problema más serio. Más bien, parecía que la definición de probabilidad, definida en términos de eventos equiprobables, necesitaba ser reconsiderada. Este hecho se expone en las discusiones hechas acerca de paradojas formuladas por el francés Joseph Bertrand y que aparecen en su *Calcul des Probabilités* de 1889.

2.4.1 La paradoja de las tres cajas de joyas

En esta paradoja consideramos tres cajas, cada una de ellas tiene dos cajones. Los dos cajones de la caja A guardan adentro una moneda de oro, los dos cajones de la caja B guardan una moneda de plata, y la caja C guarda una moneda de oro en un cajón y una moneda de plata en el otro.

Elegimos una caja al azar. La probabilidad de que elegimos la caja C es de $1/3$.

Supongamos que abrimos al azar uno de los cajones de la caja que hemos elegido. Entonces hay dos casos:

- Encontramos una medalla de oro. En este caso quedan sólo dos posibilidades: el otro cajón contiene una moneda de oro (elegimos la caja A) o el otro cajón tiene una moneda de plata (elegimos caja C).

- Encontramos una moneda de plata. En este caso también quedan sólo dos posibilidades: el otro cajón tiene una moneda de oro (elegimos caja C) o el otro cajón tiene una moneda de plata (elegimos caja B).

Se tiene que, cualquiera que sea el caso, quedan ahora dos posibilidades, una de las cuales es que hemos elegido la caja C. Por lo tanto, la probabilidad de que hemos elegido la caja C es ahora $1/2$.

Bertrand mismo no aceptaba la conclusión de que abrir un cajón cambiara la probabilidad de elegir la caja C de $1/3$ a $1/2$. Poincaré (1912, páginas 26-27) dio la siguiente explicación: supongamos que los cajones están etiquetados (en un lugar donde no podemos ver) como α y β, y supongamos que la moneda de oro en la caja C esta en el cajón α. Entonces hay 6 casos iguales para la caja que abrimos:

1. Caja A, cajón α: moneda de oro.
2. Caja A, cajón β: moneda de oro.
3. Caja B, cajón α: moneda de plata.
4. Caja B, cajón β: moneda de plata.
5. Caja C, cajón α: moneda de oro.
6. Caja C, cajón β: moneda de plata.

Si al abrir una caja encontramos una moneda de oro, entonces tres de estos casos permanecen posibles: caso 1, caso 2 y caso 5. De los tres, sólo uno favorece el haber elegido la caja C, de modo que la probabilidad para la caja C es todavía $1/3$.

El siguiente principio surge en la solución de este problema: Los casos equiprobables deben estar descritos con suficiente detalle, de manera que representen información nueva (por ejemplo, al saber que se encontró una moneda de oro) con todo detalle relevante.

Un problema relacionado con esta paradoja es el *Problema de Monty Hall*, sobre un juego donde se tienen tres puertas que esconden tras de sí dos cabras y un carro. El jugador elige una puerta. El presentador abre una de las dos puertas que no eligió el jugador, detrás de la cual aparece una cabra. Se le permite de nuevo al jugador elegir de entre las dos puertas sin abrir. ¿Debería el jugador cambiar de elección?

2.4.2 La paradoja del círculo máximo

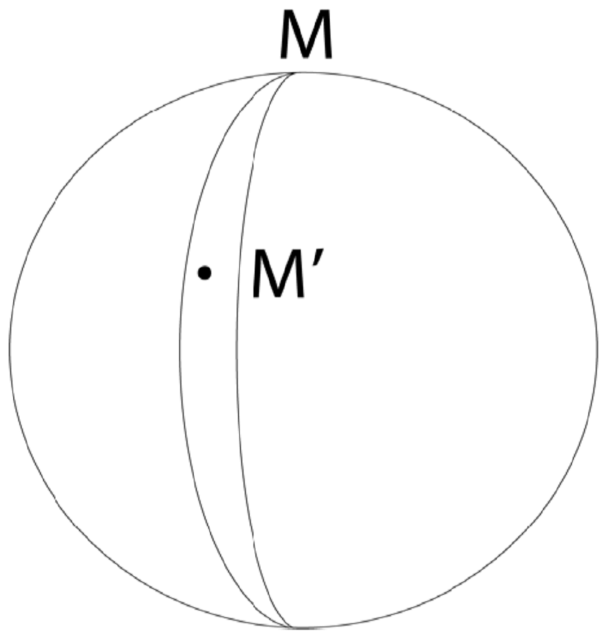
En esta paradoja, Bertrand empieza con la pregunta: ¿cuál es la probabilidad de que al escoger dos puntos al azar en la superficie de una esfera la distancia entre ellos sea menor que 10 min?

Por la simetría de la esfera, podemos suponer que el primer punto es conocido. Así, una manera de responder a la pregunta es calculando la proporción de la superficie de la esfera que se halla a una distancia de 10 min del punto dado, esto es 2.1×10^{-6}.

Bertrand también dio otra solución a este problema. Después de fijar el primer punto, también podemos suponer que conocemos el círculo máximo que conecta los dos puntos, ya que las posibilidades de que el segundo punto caiga en uno u otro círculo máximo en torno al primer punto son las mismas. Hay 360 grados, esto es, 2160 arcos de 10 min, en este círculo máximo. Solamente los puntos en los dos arcos vecinos al primer punto se hallan a una distancia de a lo más 10 min, de manera que la probabilidad que buscamos es de $2/2160$, o bien, 9.3×10^{-4}.

Borel dio la siguiente explicación sobre esta paradoja en su libro de texto de probabilidad publicado en 1909. Él afirmó que el primer método utilizado por Bertrand, en donde hace la suposición de que áreas iguales tienen probabilidades iguales de contener al segundo punto, es correcto. Su segundo método, basado en el supuesto de que arcos iguales en el círculo máximo tienen probabilidades iguales de contener al segundo punto, es incorrecto. Escribiendo M y M$'$ para los puntos que serán elegidos al azar en la esfera, Borel explicó el error de Bertrand de la siguiente manera:

> ...El error empieza cuando, después de fijar el punto M y el círculo máximo, uno supone que la probabilidad de que M$'$ se halle en un arco dado del círculo máximo es proporcional a la longitud de ese arco. Si el arco no tiene grosor, entonces, hablando en el sentido estricto, habremos de asignar el valor cero a la probabilidad de que M y M' se hallen en el mismo círculo. Con el fin de evitar este factor de cero, lo cual hace imposible cualquier cálculo, debemos de considerar franjas delgadas de círculos máximos, los cuales pasan todos por M, donde las franjas son de tal manera que éstas se hacen cada vez más delgadas al acercarse a los polos -de los cuales M es uno de ellos- de modo que las franjas no se traslapen una sobre la otra. (Véase la figura).



De la solución a esta paradoja surge el siguiente principio: Es posible que necesitemos considerar un evento real observado de probabilidad distinta de cero como uno que es representado de una manera idealizada por un evento de probabilidad cero.

La confusión en torno a las paradojas significó una fuente más de insatisfacción con la teoría clásica, basada en eventos equiprobables.

3 David Hilbert y la Axiomatización de la Probabilidad

A inicios del siglo XX, muchos matemáticos se hallaban insatisfechos con aquello que ellos veían como una falta de claridad y rigor en el cálculo de las probabilidades. Todos estos cálculos parecían depender por completo de conceptos que reposan fuera de las matemáticas: evento, ensayo, aleatoriedad, probabilidad. Como Henri Poincaré escribió: "difícilmente puede uno dar una definición satisfactoria de probabilidad" (Poincaré, 1912).

3.1 El Sexto Problema de Hilbert

El llamado más célebre por una aclaración sobre el tratamiento de la probabilidad vino de David Hilbert. Al respecto, mencionamos la versión escrita del Sexto de los Veintitrés Problemas que Hilbert planteó en el Congreso Internacional de Matemáticas en París en 1900, que a la letra dice:

> La investigación de los fundamentos de la geometría sugiere el problema: Considerar de la misma forma, por medio de axiomas, aquellas ciencias físicas en las cuales las matemáticas desempeñan una parte importante; en el primer nivel se hallan la teoría de las probabilidades y la mecánica (Hilbert, 1902).

Para dar a entender a qué se refería por axiomas de probabilidad, Hilbert citó como ejemplo a Georg Bohlmann, quien en 1901, había asignado a las reglas de probabilidad total y probabilidad compuesta la categoría de axiomas, más bien que de teoremas. Específicamente, para los eventos E_1, y E_2, Bohlmann estableció los axiomas

$$\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2), \quad \text{si } \mathbb{P}(E_1 \cap E_2) = 0.$$

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) \cdot \mathbb{P}(E_1 \mid E_2).$$

3.2 *Grundlagen der Geometrie*

En *Grundlagen der Geometrie* (*Fundamentos de la Geometría*), publicado en 1899, Hilbert sustituye los axiomas de Euclides tradicionales por un conjunto formal de 21 axiomas. En este texto Hilbert expone su enfoque axiomático.

El enfoque de Hilbert marcó el cambio al sistema axiomático moderno. Los axiomas no se toman como verdades evidentes. La geometría puede tratar de cosas, sobre las que tenemos intuiciones poderosas, pero no es necesario asignar un significado explícito a los conceptos indefinidos. Como dice Hilbert, los elementos tales como el punto, la recta, el plano y otros, se pueden sustituir con mesas, sillas, jarras de cerveza y otros objetos. Lo que se discute son sus relaciones definidas. Por esta razón, a menudo se dice que Hilbert promovió un punto de vista en el que las matemáticas es vista como un juego vacío formal en el que las reglas para hacer deducciones se dan por adelantado.

En Corry (1997) podemos encontrar más acerca del método axiomático de Hilbert.

3.3 Algunas Respuestas al Sexto Problema

Entre las primeras respuestas al desafío que Hilbert planteaba podemos mencionar las siguientes.

En 1904 Rudolf Laemmel, en su tesis doctoral escrita en Zurich, propone dos axiomas y tres definiciones como un "sistema mínimo" para la teoría, formulándolos en términos de nociones de la teoría de conjuntos. Luego afirmó que sus axiomas eran independientes y suficientes para desarrollar toda la teoría, pero no mencionó el problema de la consistencia de los axiomas.

En 1907, Ugo Broggi, uno de los estudiantes de doctorado de Hilbert hizo un intento de perfeccionar las propuestas dadas antes por Bohlmann y Laemmel. Basándose en la teoría de la medida de Lebesgue, Broggi no solamente formuló un sistema de axiomas para la probabilidad, sino que también mostró que sus axiomas estaban completos (en el sentido de Hilbert), independientes y consistentes, demostrando así las deficiencias del sistema previo propuesto por Bohlmann. En 1908, en un discurso pronunciado en el 4to Congreso Internacional de Matemáticas en Roma, Bohlmann hizo referencia al trabajo doctoral de Broggi y reconoció que en este trabajo se dejaba ver la necesidad de proporcionar un análisis lógico mas profundo del concepto de evento dentro de la teoría de probabilidades.

En un artículo publicado en ruso, Sergei Bernstein (1917) propuso que la teoría de la probabilidad se puede basar sobre axiomas cualitativos para coeficientes numéricos que miden la probabilidad de una proposición. También desarrolló esta idea en su libro de texto de probabilidad (Bernstein, 1927), y Kolmogorov mencionó ambos, el articulo y el libro, en la bibliografía de *Fundamentos.*

4 La Teoría de la Medida Antes de los Axiomas de Probabilidad

Entre los trabajos principales en el trayecto que llevó hacia *Fundamentos* de Kolmogorov se hallan los llevados a cabo por Émile Borel quien planteó y dio solución a problemas relativos a eventos cuya ocurrencia dependiera de una infinidad de ensayos, logrando con esto rebasar el marco clásico del estudio de la probabilidad.

Fueron muchos los matemáticos quienes hicieron importantes contribuciones en el estudio de la teoría de la medida y de su utilidad en el estudio de la teoría de la probabilidad. A continuación se hace una breve mención de algunos trabajos de estos matemáticos, la mayoría de los cuales fueron publicados entre los años 1900-1930. Entre los resultados que destacan se hayan aquellos sobre sumas de variables aleatorias independientes: la Ley de los Grandes Números, el Teorema del Límite Central y la Ley del Logaritmo Iterado. La exposición está ordenada por escuelas.

4.1 La Escuela Francesa

Las bases para la transformación de la teoría de la probabilidad característica del siglo XX fueron dadas por **Henri Lebesgue** (1847-1941) en su tesis (Lebesgue, 1902) *Intégrale, Longeur, Aire* (*Integral, Longitud, Área*). Su contribución clave fue el uso sistemático de funciones contablemente aditivas $\mu(.)$, de manera que

$$\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$$

para conjuntos ajenos A_n de alguna clase apropiada, a saber, la σ-álgebra de conjuntos medibles.

Maurice Fréchet (1878-1973) introdujo en su tesis (Fréchet, 1906), el concepto ahora tan utilizado de espacios métricos. La obra de Fréchet inicia la tendencia hacia la abstracción y la axiomatización características del análisis moderno. En particular, condujo al desprendimiento de la teoría de la medida de Lebesgue de sus orígenes Euclídeos, por ejemplo, en lugar de considerar sólo funciones cuyo dominio son subconjuntos de $\mathbb{R}^n$, también podemos considerar funciones cuyo dominio sea el de subconjuntos del espacio relacionado a todos los resultados posibles que se obtienen de una sucesión infinita de lanzamientos de una moneda.

Émile Borel (1871-1956), con su alumno Lebesgue, publicó en 1909 *Les Probabilités Dénombrables et leurs Applications Arithmétiques* (*Las Probabilidades Numerables y sus Aplicaciones Aritméticas*). En esta obra Borel hizo una clasificación en dos categorías de los problemas de probabilidad estudiados en el marco de la teoría clásica: en la primera categoría están las probabilidades discontinuas, en donde el número de casos posibles es finito. En la segunda categoría están las probabilidades geométricas o probabilidades en el dominio continuo, en las cuales el número de casos es infinito.

Valiéndose de resultados de la teoría de conjuntos, Borel encontró que tal clasificación aparecía como incompleta. Hacía falta estudiar problemas en los que el número de resultados posibles fuera infinito numerable. Borel emprendió el estudio de problemas que caen en esta categoría. A las cuestiones relacionadas con este tipo de problemas Borel las llamó *Probabilidades Numerables*.

La obra de Borel es destacable por varios resultados importantes que aparecen en ella. Algunos de estos resultados son el Teorema de Números Normales de Borel y el lema de Borel-Cantelli. El corolario del Teorema de Números Normales, que tales números existen, es llamativo, ¡no se conoce ningún ejemplo específico!.

Paul Lévy (1886-1971), uno de los grandes probabilistas del siglo, escribió *Calculs des Probabilités* (Lévy, 1925), un libro de texto de probabilidad basado en ideas de teoría de la medida en donde introduce y estudia las funciones características. Asimismo creó la teoría de procesos estocásticos con 'incrementos estacionarios independientes', ahora conocidos como procesos de Lévy.

4.2 La Escuela Polaca

En el período que abarca entre la primera y la segunda guerra mundial, hubo en Polonia un gran crecimiento en el área de las matemáticas, y particularmente, en el área de análisis. Varios Matemáticos polacos de este período trabajaron en probabilidad, empezando con **Hugo Steinhaus** (1887-1972), quien en 1923 dio un intento de axiomatización (Steinhauss, 1923). Entre la escuela polaca también se encuentran las colaboraciones importantes de Jozef Marcinkiewicks y Antoni Zygmund. El trabajo polaco se centró en las llamadas funciones independientes como un puente entre probabilidad y análisis.

Stefan Banach (1892-1945), en su libro (Banach, 1932) demostró muchos teoremas fundamentales del análisis funcional y formuló el concepto ahora conocido como Espacios de Banach.

4.3 La Escuela Rusa

La escuela Rusa logra de manera exitosa tomar los resultados de la teoría de la medida y aplicar éstos a la teoría de probabilidad.

Sergei Natanovich Bernstein (1880-1968), maestro en Kharkov, escribió ampliamente sobre probabilidad y análisis, incluyendo los polinomios de Bernstein y su aplicación a la ley de los grandes números (Bernstein, 1911).

Alexander Yakovlevich Khinchine (1894-1956). Su más grande logro es, podría decirse, su ley de logaritmo iterado de 1924, completando así, junto con la ley de los grandes números y el teorema del límite central el trío clásico de teoremas límite en la teoría de la probabilidad.

Andrei Nikolaevich Kolmogorov (1903-1987) escribió alguna docena de artículos sobre probabilidad entre su primer trabajo en esta área, con Khinchine en 1925, arriba mencionado, y su trascendental obra *Fundamentos* (Kolmogorov, 1933). Acerca de este trabajo se discute más adelante.

Andrey Andreyevich Markov (1856-1922) es mayormente conocido por su trabajo en teoría de procesos estocásticos, en particular por su estudio de las cadenas de Markov, sucesiones de variables aleatorias en las que los futuros valores de la variable dependen del valor presente de la variable, pero no dependen de sus valores anteriores.

Aleksandr Mikhailovich Lyapunov (1857-1918) junto con Andrey Markov, fue estudiante de Chebyshev. Lyapunov probó el teorema del límite central usando variables aleatorias independientes, no necesariamente idénticamente distribuidas, a diferencia de la prueba clásica de este teorema, la cual requiere de variables aleatorias independientes e idénticamente distribuidas.

4.4 La Escuela Italiana

Francesco Paolo Cantelli (1906-1985) es recordado por dos contribuciones principales. La primera, de 1917, es su parte en los lemas de Borel-Cantelli. La segunda, de 1933, es su trabajo en el teorema de Glivenko-Cantelli o teorema Fundamental de la Estadística, que la muestra determina, en el límite, a la población.

Bruno de Finetti (1906-1985) contribuyó a la teoría de la probabilidad con su trabajo en procesos estocásticos con incrementos independientes, o procesos de Lévy en terminología moderna, y la fórmula de Lévy-Khinchine para tales procesos, en 1929. También estudió aquellas variables aleatorias para las cuales su distribución conjunta es simétrica. Quizás el trabajo más importante de de Finetti fue su temprano apoyo al enfoque Bayesiano de la estadística, culminando en su libro (de Finetti, 1970).

4.5 La Escuela Alemana

David Hilbert, en un pie de página al sexto problema de su lista de 1900, hace referencia al trabajo de Georg Bohlmann sobre la axiomatización de la probabilidad. En Bohlmann (1908) se da una definición rigurosa de independencia, a saber, la regla del producto para independencia que previamente había sido visto como un teorema.

Felix Hausdorff (1868-1942) contribuyó a la teoría de la probabilidad en su trabajo de 1913, en el contexto que más adelante llegó a ser la ley del logaritmo iterado establecido por Khinchine en 1924. Hausdorff también dio la primera prueba rigurosa de la ley fuerte de los grandes números para ensayos de Bernoulli (Hausdorff, 1914).

Richard von Mises (1833-1953), introdujo una teoría en la que se considera a Ley Fuerte de los Grandes Números como axioma, y no como teorema. Su contribución fue hacer notar que esto podía servir como base para la teoría de la probabilidad (von Mises, 1928). Lévy, entre otros, encontraron esta teoría poco satisfactoria.

George Pólya era húngaro pero sus publicaciones fueron en alemán. En Pólya (1920) introdujo el término 'Teorema del Límite Central' en alemán.

4.6 La Escuela Estadounidense

Norbert Wiener (1884-1964) escribió sobre probabilidad en una serie de artículos entre 1920 y 1923. Wiener fue el primero en construir un modelo matemático riguroso para el fenómeno físico del movimiento Browniano. Es en reconocimiento a esta labor que a menudo se refiere al movimiento Browniano como procesos de Wiener.

Willam Feller (1906-1970), nacido en Yugoslavia llegó a ser uno de tantos matemáticos prominentes que dejaron Europa y se fueron a EE.UU. durante el período Nazi. Originalmente un analista, su interés en la teoría de la probabilidad data desde su tiempo en Keil (1928-33); sus trabajos publicados sobre probabilidad datan de 1935. Sobre Feller, Doob escribió:

> "Feller hizo contribuciones profundas y originales a la teoría de la probabilidad por un período (de 1935 hasta su muerte) durante el cual la probabilidad pasó de ser una teoría con una relación pobre con las matemáticas a ser una rama central de ésta" (Doob, 1972).

Un estímulo importante para el trabajo en probabilidad y estadística en EE.UU. fue la creación en 1930 de la revista *Annals of Mathamatical Statistics.* Entre sus publicaciones tempranas más importantes podemos encontrar el trabajo de J. L. Doob (1910-2004).

4.7 La Escuela Inglesa

El trabajo de la escuela Inglesa tiene raíces en parte estadísticas y en parte filosóficas, lo cual forma un contraste con las escuelas Rusa, Francesa y Polaca cuyo trabajo en probabilidad fue altamente matemático.

John Maynard Keynes (1883-1946) escribió *A Treatise in Probability* (Keynes, 1921). Este es un tratado filosófico con poco contenido matemático.

R. A. Fisher (1890-1962), quien se halla entre los principales estadísticos, escribió en estadística de 1916 en adelante haciendo labor pionera en trabajos sobre verosimilitud y suficiencia, estimación de parámetros, análisis de varianza y diseño de experimentos.

El primer libro de texto en sintetizar exitosamente las ideas de Fisher y que esencialmente marcó el inicio de la estadística moderna es Cramér (1946). En este libro Cramér incluye los axiomas de Kolmogorov, aunque en realidad apela muy poco a ellos en el resto de su obra.

5 *FUNDAMENTOS*

En su libro clásico *Grundbegriffe der Wahrscheinlichkeitsrechnung* (*Fundamentos de la Teoría de la Probabilidad*), Kolmogorov (1933) inauguró la era moderna de la teoría de la probabilidad. Es en este libro donde Kolmogorov logra exitosamente colocar a la teoría de la probabilidad sobre una base axiomática rigurosa, aprovechando las herramientas que la teoría de la medida brindaba, viendo a la probabilidad como una medida positiva de masa uno. En particular, Kolmogorov completa con esto la solución al Problema Seis de Hilbert, sobre la axiomatización de la probabilidad.

5.1 Los Seis Axiomas

Kolmogorov comenzó con cinco axiomas acerca de un conjunto E y un conjunto $\mathcal{F}$ de subconjuntos de E, a los cuales llamó *eventos aleatorios*.

I. $\mathcal{F}$ es un álgebra de conjuntos.

II. $\mathcal{F}$ contiene al conjunto E.

III. Para cada conjunto A de $\mathcal{F}$ está asignado un número real no negativo $\mathbb{P}(A)$. Al número $\mathbb{P}(A)$ lo llamamos la probabilidad del evento A.

IV. $\mathbb{P}(E) = 1$.

V. Si A y B son ajenos, entonces

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Además agregó un sexto axioma, redundante para el caso en que $\mathcal{F}$ es finito, pero independiente de los primeros cinco axiomas para $\mathcal{F}$ infinito:

VI. Si $A_1 \supseteq A_2 \supseteq \cdots$ es una sucesión decreciente de eventos de $\mathcal{F}$ con $\cap_{n=1}^{\infty} A_n = \varnothing$, entonces $\lim_{n\to\infty} \mathbb{P}(A_n) = 0$.

Kolmogorov presenta los primeros cinco axiomas en el primer capítulo de su libro, y a su vez da la interpretación de estos axiomas en términos de probabilidad frecuencial, siguiendo las ideas de von Mises; asimismo, hace una exposición clara de los conceptos de probabilidad que pueden ser estudiados con estos axiomas elementales (independencia de eventos, probabilidad condicional, teorema de Bayes y cadenas de Markov). Este modelo matemático es el adecuado para estudiar fenómenos aleatorios con un número finito de resultados, no necesariamente con la misma probabilidad. De hecho, los axiomas I-V sirven como base axiomática para la probabilidad clásica.

En el segundo capítulo agrega el *axioma de continuidad* de la probabilidad, con lo que se establece el concepto de espacio de probabilidad. El uso de este axioma permite resolver problemas como los planteados por Borel, sobre eventos cuya ocurrencia dependen de un número infinito de ensayos.

5.2 El Axioma de Continuidad y la Aditividad Numerable

Con el fin de mostrar la relación que tienen los axiomas de probabilidad con la teoría de la medida, a continuación se demuestra la equivalencia del axioma de continuidad con la aditividad numerable. Para esto consideremos los conjuntos ajenos $A_1, A_2, ...$ de $\mathcal{F}$.

Sea
$A = A_1 \cup A_2 \cup ... \cup A_n \cup ...,$
y sea
$B_n = A_n \cup A_{n+1} \cup ....$

Entonces, $B_{n+1} \subseteq B_n$. La ocurrencia de B_n nos indica la ocurrencia de alguno de los eventos A_n, A_{n+1}, A_{n+2}, ..., digamos el $n + i$, $i = 0, 1, 2, ...$, luego del hecho de que los eventos son ajenos, tenemos que los eventos A_{n+i+1}, $A_{n+i+2}, ...$no suceden. Así, los eventos B_{n+i+1}, B_{n+i+2}, ... no suceden, y por lo tanto,

$$\cap_{n=1}^{\infty} B_n = \varnothing.$$

Del axioma de continuidad se sigue que

$$\lim_{n\to\infty} \mathbb{P}(B_n) = 0.$$

Por otra lado tenemos que $A = A_1 \cup ... \cup A_n \cup B_{n+1}$, donde A_1,..., A_n, B_{n+1} son ajenos. Por lo tanto, por el axioma III, tenemos que

$$\mathbb{P}(A) = \mathbb{P}(A_1) + ... + \mathbb{P}(A_n) + \mathbb{P}(B_{n+1})$$

$$= \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(A_k) + \lim_{n\to\infty} \mathbb{P}(B_n)$$

$$= \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

5.3 La Ley Fuerte de los Grandes Números

Hay un punto que vale la pena mencionar aquí, a saber, la forma final de la Ley Fuerte de los Grandes Números.

Teorema 3 *Sean X, X_1, $X_2, ...$ variables aleatorias independientes, idénticamente distribuidas, sea $\mu = \mathbb{E}X < \infty$, y sea $S_n = \sum_{i=1}^{n} X_i$, entonces,*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) = 1.$$

Podemos observar que la ley fuerte de los grandes números permite "meter" en el argumento de $\mathbb{P}$ el límite que aparece en (1). Este paso requiere el axioma de continuidad. Otro hecho acerca de estas dos formas de la ley de los grandes números es que la ley fuerte implica la ley débil.

Esto completa la línea de trabajo empezada por Bernoulli en su póstumo clásico *Ars Conjectandi* de 1713, en el cual aparece el teorema de Bernoulli, o la ley débil de los grandes números para ensayos de Bernoulli. Esta es la forma precisa de la idea popular de la 'Ley de Promedios', y muestra convincentemente que los axiomas de Kolmogorov captan exitosamente la esencia de los orígenes de la probabilidad.

6 Conclusiones

La teoría de la probabilidad es una teoría matemática que se desarrolla a partir de axiomas (espacio de probabilidad). Esta teoría no depende de nociones como evento, aleatoriedad, etc., aunque son precisamente estas nociones las que motivan el estudio de esta materia.

En el logro de la axiomatización destacan los siguientes elementos: La teoría de la medida, la cual provee de herramientas matemáticas poderosas para estudiar la probabilidad, el llamado de Hilbert a la axiomatización de ésta y los axiomas de Kolmogorov, con los cuales se da solución al Sexto Problema de Hilbert.

7 Referencias

BANACH, S. (1932). Théories des Opérations Linéaires. Warsaw: Monografje Matematyczne.

BARBEAU, E. B. (2000). Mathematical Fallacies, Flaws and Flimflam (Spectrum), Edward J. Barbeau, The Mathematical Association of America, pp 77-90.

BERNOULLI, J. (1713). *The Art of Conjecturing* (El arte de predecir). Jacob Bernoulli. Traducción al inglés con introducción y notas de Edith Dudley Sylla. Johns Hopkins Press. 2006.

BERNSTEIN, S. N. (1911). Probability Theory (en ruso). St. Petesburg: Academy oy Sciences.

BERNSTEIN, S. N. (1917). On the axiomatic foundation of the theory of probability.*Communications of the Kharkiv Mathematical Society* **15** 209–274. Reprinted in S. N. Bernstein (1964). Nauka, Moscow.

BERNSTEIN, S. N. (1927). *Theory of Probability.* State Publishing House, Moscow and Leningrad. Second edition 1934, fourth 1946. Esta obra estaba incluída en la bibliografía del *Grundbegriffe.*

BERTRAND, J. (1889). Calcul des probabilités. Gauthier-Villars, Paris. Some copies of the first edition are dated 1888. Second edition 1907. Reprinted by Chelsea, New York, 1972.

BINGHAM N. H. (2000). Studies in the History of Probability and Statistics KLVI. Measure into Probability: From Lebesgue to Kolmogorov, *Biometrika*, Vol. 87, No. 1. pp. 145-156.

BOHLMANN, G. (1901). Lebensversicherungs-Mathematik. In *Encyklopädie der Mathematischen Wissenschaften* **1**(2) 852–917. Teubner, Leipzig.

BOHLMANN, G. (1908) Die Grundbegriffe der Wahrscheinlichkeitsrechnung in ihrer Anwendung auf die Lebensversicherung, in G. CASTELNUOVO (ed.) *Atti del IV congresso internazionale dei mathematici* (1908), *Roma, Academia dei Lincei*, 244-278.

BOREL, E. (1909). Les probabilités dénombrables et leur applications aritmétiques. *Rend. Circ. Mat. Parlermo* **27**, 247-71.

BROGGI, U. (1907). Die Axiome der Wahrscheinlichkeitsrechnung. Ph.D. thesis, Universität Göttingen. Excerpts reprinted in Schneider (1988) 359–366.

CORRY, L (1997). David Hilbert and the axiomatization of physics. Arch. Hist. Exact Sci **51**, 83-198.

CRAMÉR, H. (1946). Mathematical Methods of Statistics. Princeton: Princeton Univ. Press.

DOOB, J. L. (1972). William Feller and twentieth century probability. Proc. 6th Berkeley Symp. Math Statist. Prob., 2, pp. xv-xx. Berkeley, CA: Univ. California Press (Math. Rev. 52 #7795).

FINETTI, B. DE (1970). Teoria delle probabilità: sistesi introductiva con apéndice critica. Turín: Nuova Bibl. Sci. Einaudi, G. Einaudi Editore (trans. 1974, 1975 as Theory of probability: A Critical Introduction, **1**, **2**. London: Wiley).

FRÉCHET, M. (1906). Sur quelques points du calcul fonctionnel. *Red. Cir. Mat. Palermo* **22**, 1-74.

FRÉCHET, M. (1951). Rapport général sur les travaux du Colloque de Calcul des Probabilités. In Bayer (1951) 3–21.

GLENN S. and VLADIMIR V. (2006) The Sources of Kolmogorov's Grundbegriffe, Statistical Science, Vol. 21,No. 1, 70-98.

GNEDENKO B. V. Theory of Probability, Boris V. Gnedenko, 6^a Edición, Gordon and Breach Science Publishers, pp 26-34.

HAUSDORFF, F. (1914) Grundzüge der Mengenlehre, Leipzig (3rd ed. 1935, Berlin, reprinted 1944, New York: Dover).

HILBERT, D.(1899) Grundlagen der Geometrie (Festschrift zur Feier der Enthiillung des Gauss-Weber- Denkmals in GiSttingen), Leipzig, Teubner.

HILBERT, D. (1902). Mathematical problems. Bull. Amer. Math. Soc. **8** 437–479. Traducido del alemá por M. W. Newson.

KEYNES, J. M. (1921). A Treatise on Probability. Macmillan, London.

KOLMOGOROV, A. N. (1933). Foundations of Probability Theory. Traducción al inglés de N. Morrison. Chelsea, New York. 1950

LAEMMEL, R. (1904). Untersuchungen über die Ermittlung von Wahrscheinlichkeiten. Ph.D. thesis, Universität Zürich. Excerpts reprinted in Schneider (1988) 367–377.

LEBESGUE, H. (1902). Intégrale, longueur, aire. *Annali di Mat.* (**3**) 7, 231-59.

LÉVY, P. (1925). *Calcul des probabilités.* Gauthier-Villars, Paris.

MARTIN, T. (2003). Probabilité et certitude. In Probabilités subjectives et rationalité de l'action (T. Martin, ed.) 119–134. CNRS Éditions, Paris.

de MOIVRE, A. (1718). The Doctrine of Chances: Or, A Method of Calculating the Probability of Events in Play. Pearson, London. Second edition 1738, third 1756.

POINCARÉ, H. (1912). *Calcul des probabilités.* Gauthier-Villars, Paris. Second edition of Poincaré (1896).

PÓLYA, G. (1920).Über den zentrales Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. Math. Z. 8. 171-81.

STEINHAUS, H. (1923). Les probabilités dénombrables et leur rapport a la théorie de la mesure. Fund. Math. 4 286–310.

VENN, J. (1888). The Logic of Chance, 3rd ed. Macmillan, London and New York. First edition 1866, second 1876.

VON MISES, R. (1928) Wahrscheinlichkeit, statistic und Wahrheit (3rd ed. 1951; trans. 1939, reprinted 1981, as Probability, Statistics and Truth, New York: Dover)