TEORÍA DE LA PROBABILIDAD

Desarrollo histórico de la Teoría de la Probabilidad

El Cálculo de Probabilidades tuvo su origen en el estudio de algunos problemas relacionados con juegos de dados. La dificultad y al mismo tiempo lo interesante de estos juegos consiste en que al lanzar uno o varios dados, de la manera usual en que esto se realiza, el resultado que se obtiene no está únicamente determinado, sino que es uno de un conjunto de posibles resultados.

En situaciones como la del lanzamiento de varios dados, donde existen diferentes posibles resultados, el resultado es únicamente uno de ellos, pero, a priori, no se sabe cuál; de ahí que podemos hacer aseveraciones acerca del resultado que obtendremos, las cuales tienen la característica de que pueden resultar ser verdaderas o falsas una vez que observamos el resultado obtenido. Por ejemplo, al lanzar tres dados, la aseveración "la suma de los números que muestran las caras superiores de los dados, una vez lanzados, es mayor que 1" puede resultar verdadera o falsa cuando se realiza el lanzamiento. A cada una de esas aseveraciones que se pueden hacer acerca del resultado que se obtiene se le llama un **evento**; cuando la aseveración resulta ser verdadera, decimos que el evento ocurre; cuando resulta falsa, decimos que no ocurre.

Considerando que un evento puede ocurrir o no ocurrir, lo que se busca es asignar un número real no negativo a cada evento, el cual indique que tanto se puede esperar que el evento ocurra. A ese número se le llama la probabilidad del evento. De esta forma, la probabilidad la podemos ver como una función no negativa definida sobre la familia de eventos, a la cual llamaremos **función de probabilidad**.

Para fines de aplicar el Cálculo de Probabilidades se introduce el concepto de **experimento aleatorio**, donde un experimento se considera como cualquier proceso que conduzca a un resultado. Algunos experimentos pueden consistir simplemente en la observación de un determinado fenómeno natural y en ir anotando algunas de las características que observamos; por ejemplo, la medición día con día del crecimiento de una planta. En otros experimentos están involucradas acciones nuestras, además de la observación, como en el caso del lanzamiento de un dado.

Un experimento consta de dos partes, por un lado tenemos la descripción del proceso que estamos considerando, especificando las condiciones bajo las cuales se realiza, y por otro lado tenemos lo que consideramos como su resultado.

Se dice que un experimento es aleatorio cuando al realizarse, bajo las condiciones que se indiquen, su resultado puede ser cualquiera de un conjunto de resultados posibles. Cabe aclarar que la posibilidad de distintos resultados puede provenir de que las condiciones que se especifican para la realización del experimento incluyen cierta arbitrariedad.

En la formulación moderna, para modelar matematicamente un experimento aleatorio se busca construir un espacio de medida $(\Omega, \Im, P)$, donde Ω, llamado el espacio muestral del experimento, es el conjunto de sus posibles resultados, $\Im$ es una σ-álgebra de subconjuntos de Ω, los cuales representan a los eventos, y la medida P, definida sobre $\Im$, tiene la característica de asignarle a Ω el valor 1. Cada elemento de $\Im$ es entonces un subconjunto del total de posibles resultados del experimento aleatorio en consideración y lo que nos interesa obtener es la probabilidad de que ocurra alguno de los posibles resultados que pertenecen a ese subconjunto. En general, la medida P se construye mediante un proceso de extensión: se comienza por asignar probabilidades a algunos eventos y después, utilizando las propiedades de P, se busca extenderla hasta abarcar el mayor número posible de subconjuntos de Ω. De esta forma, el modelo matemático para el estudio del experimento consta de 3 elementos, el primero es el conjunto Ω, el segundo una familia de subconjuntos de Ω y el tercero una medida, que llamaremos medida de probabilidad, la cual asocia a cada elemento A de $\Im$ un número real que designa la probabilidad de que ocurra alguno de los posibles resultados que pertenecen a A.

Este modelo se fue construyendo durante los primeros 33 años del siglo XX, buscando axiomatizar el Cálculo de Probabilidades. Durante esos 33 años se estableció y consolidó su vínculo con la Teoría de la Medida, la cual surgió en los primeros años del siglo con los trabajos de Émile Borel y Henri Lebesgue. **Es con su formulación axiomática que podemos hablar de una teoría matemática de la probabilidad, ya que de esta forma se tiene un cuerpo teórico, independiente de los problemas reales específicos que se tratan con el Cálculo de Probabilidades**.

La identificación de una función de probabilidad con una medida no surgió de manera automática como algo general aplicable a cualquier fenómeno aleatorio, sino que requirió de un proceso, que llevó varios años, en el cual se mostró que es una alternativa adecuada. En el centro de este proceso se encuentra el planteamiento de problemas donde se trata de calcular probabilidades de eventos cuya ocurrencia o no ocurrencia depende de una infinidad de observaciones y la aceptación de la σ-aditividad como una propiedad de cualquier función de probabilidad, la cual permite atacar ese tipo de problemas.

Origen del Cálculo de Probabilidades

Los juegos donde se utilizan dados o algo similar se inventaron hace miles de años; el dado más antiguo que se conoce fue encontrado en Irak y tiene una antigüedad de alrededor de 5 mil años. El astragalo, también conocido como tali o taba, es un precursor del dado. Se trata de un hueso del pie de un animal, el cual presenta cuatro caras, cada una con un nombre que más tarde se convirtió en un número. Las más utilizadas son las del carnero. Los juegos con tabas son tan antiguos como los juegos con dados; se han encontrado tabas en las tumbas de los faraones egipcios. Incluso hoy en día hay lugares, incluyendo algunas regiones de México, donde se juega con una taba.

En la antigua Grecia y en Roma se volvió muy popular un juego que consistía en lanzar 4 tabas; en las caras de cada una de ellas estaban marcados los números 1, 3, 4 y 6, respectivamente. Las diferentes combinaciones tenían diferentes valores, siendo considerada la más alta la que consiste en 4 números distintos, la cual era llamada Venus.

Los aztecas y otros pueblos originarios de América utilizaban algo que puede considerarse equivalente a los dados; se trata de unos granos rojos parecidos al frijol, cada uno de los cuales se pintaba de un lado. Esos granos son semillas de un árbol que era conocido como tzité y a los granos mismos se les llamaba tzité. Actualmente se les conoce como colorines. Los granos eran utilizados para fines adivinatorios y para jugar un juego que llamaban patolli, el cual se jugaba sobre un tablero en forma de cruz diagonal, dividida en 52 casillas y se utilizaban 5 colorines marcados cada uno de un lado. Cada jugador colocaba una apuesta (había quienes apostaban incluso su persona y si perdían quedaban sometidos a la condición de esclavos). Ganaba el juego quien lograra primero recorrer las 52 casillas, avanzando según el número de caras marcadas que se obtienen al lanzar los 5 colorines.

Durante la Edad media los juegos con dados se extendieron por toda Europa y se fue adquiriendo un conocimiento empírico acerca de ellos. Fue en esa época cuando se acuñó el término azar, el cual es de origen árabe y surgió a fines del siglo XI. En una de las caras de los dados que se utilizaban para jugar, estaba dibujada una flor, representando un resultado desfavorable para quien lo obtenía. La expresión árabe es az-zahr, la cual significa la flor. Por extensión se llamaba también az-zahr al dado y también se entendía por az-zahr el lanzar los dados.

La experiencia fue indicando que algunas combinaciones que se obtenían al lanzar los dados eran menos frecuentes que otras y quienes jugaban las conocían. Sin embargo, hasta la Edad Media, nadie intentó desarrollar una teoría matemática de los juegos con dados. No fue sino hasta el Renacimiento cuando se comenzó a teorizar acerca de ese tema.

El primer estudio sistemático de problemas relacionados con el lanzamiento de varios dados y los diferentes resultados que pueden obtenerse lo realizó Gerolamo Cardano (1501-1576), en su libro *Liber de ludo aleae* (Libro de los juegos de azar), sin embargo su trabajo tuvo poca influencia, pues si bien fue escrito en 1564, se publicó en 1663.

Fue en el año 1654 cuando se plantearon algunos problemas cuyas soluciones condujeron al establecimiento de reglas generales para calcular probabilidades. Pareciera, por la referencias que hay, que en esa época los juegos con dados eran muy populares y se conocían algunas reglas que permitían a los jugadores conocer, de manera aproximada y siempre con incertidumbre, qué posibilidades tenían de ganar al realizar una apuesta. Antoine Gombaud (1607-1684), conocido como el chevalier de Méré, jugó un papel importante para que se desarrollara un Cálculo de Probabilidades, no por aportaciones que hubiera hecho, sino porque fue él quien, en el año 1654, le planteó a Pascal dos problemas que, al ser resueltos de manera independiente por Pascal y Fermat en el mismo año, y por Huygens 3 años más tarde, marcaron la pauta para poder dar solución a una diversidad de problemas de probabilidad, lo cual condujo a una formulación general de las reglas que se requerían para el desarrollo de un Cálculo de Probabilidades.

Era del conocimiento del Chevalier de Méré que al lanzar un dado 4 veces consecutivas, había más posibilidades de obtener el número 6 por lo menos en uno de los lanzamientos que no obtener ninguno: sin embargo tenía duda acerca de cuántos lanzamientos de un par de dados se requieren para que haya más posibilidades de obtener par de seises por lo menos en uno de los lanzamientos del par de dados, que no obtener ninguno. Pensaba que la solución era tal vez 24 lanzamientos ya que argumentaba: Al lanzar un dado se tienen 6 posibles resultados y en ese caso se requieren 4 lanzamientos para apostar, con ventaja, que se obtiene el número 6 por lo menos en una ocasión; al lanzar un par de dados se tienen 36 posibles resultados y como 6 es a 4 como 36 a 24, entonces se requieren 24 lanzamientos del par de dados para apostar, con ventaja, que se obtiene par de seises por lo menos en una ocasión.

Al tener duda acerca de este resultado, le planteó el problema a Pascal, quien encontró que se requieren no 24, sino 25 lanzamientos del par de dados para que sea más favorable obtener por lo menos un par de seises que no obtener ninguno. Pascal a su vez le planteó el problema a Fermat, quien, sin conocer el método empleado por Pascal, llegó a la misma conclusión.

Se desconocen los métodos que utilizaron Pascal y Fermat para resolver el problema de los dados planteado por de Mére; sin embargo, en una carta dirigida a Pascal, Fermat desarrolló la solución de un problema con dados y de ahí podría inferirse que el método utilizado por Fermat para resolver el problema del lanzamiento de un par de dados fue como sigue:

Si trato de hacer un par de seises en n lanzamientos de un par de dados y, después que el dinero está en juego, convenimos en que no haré el primer lanzamiento, entonces es necesario que saque del juego $\frac{1}{36}$ del total, el cual denotaremos por A. Si después de eso, convenimos en que no haré el segundo lanzamiento, entonces debo sacar $\frac{1}{36}$ de lo restante, es decir de $A - \frac{1}{36}A = \frac{35}{36}A$, con lo cual se obtiene $\frac{35}{1296}A$; y si después de eso convenimos en que no haré el tercer lanzamiento, entonces debo sacar $\frac{1}{36}$ de lo restante, es decir de $\frac{35}{36}A - \frac{35}{1296}A = \frac{1225}{1296}A$, con lo cual se obtiene $\frac{1225}{46656}A$; si todavía se conviene en que no haga el cuarto lanzamiento debo sacar $\frac{1}{36}$ de lo restante, es decir de $\frac{1225}{1296}A - \frac{1225}{46656}A = \frac{42875}{46656}A$, con lo cual se obtiene $\frac{42875}{1679616}A$. Este proceso continuaría hasta que la suma de lo que le corresponde al jugador, desde el primer paso hasta el último, sea mayor que $\frac{1}{2}A$.

Como puede verse, así planteada, la solución de Fermat parece demasiado laboriosa, con números demasiado grandes; sin embargo, se puede simplificar el proceso de una manera que parece obvia, pero se desconoce si Fermat lo hizo. En efecto, el razonamiento anterior se puede escribir de la siguiente manera:

Si trato de hacer un par de seises en n lanzamientos de un par de dados y, después que el dinero está en juego, convenimos en que no haré el primer lanzamiento, entonces es necesario que saque del juego $\frac{1}{36}$ del total, el cual denotaremos por A. Si después de eso, convenimos en que no haré el segundo lanzamiento, entonces debo sacar $\frac{1}{36}$ de lo restante, es decir de $A - \frac{1}{36}A = \frac{35}{36}A$, con lo cual se obtiene $\frac{35}{(36)^2}A$; y si después de eso convenimos en que no haré el tercer lanzamiento, entonces debo sacar $\frac{1}{36}$ de lo restante, es decir de $\frac{35}{36}A - \frac{35}{(36)^2}A = \frac{(35)^2}{(36)^2}A$, con lo cual se obtiene $\frac{(35)^2}{(36)^3}A$. si todavía se conviene en que no haga el cuarto lanzamiento debo sacar $\frac{1}{36}$ de lo restante, es decir de $\frac{(35)^2}{(36)^2}A - \frac{(35)^2}{(36)^3}A = \frac{(35)^3}{(36)^3}A$, con lo cual se obtiene $\frac{(35)^3}{(36)^4}A$. De aquí ya puede verse que, en el n-simo paso, le corresponde al jugador $\frac{(35)^{n-1}}{(36)^n}$. Por lo tanto, la no realización de los n lanzamientos le vale al jugador la suma:

$$\frac{1}{36}A + \frac{35}{(36)^2}A + \frac{(35)^2}{(36)^3}A + \frac{(35)^3}{(36)^4}A + \cdots + \frac{(35)^{n-1}}{(36)^n}A = \frac{\frac{1}{36} - \frac{(35)^n}{(36)^{n+1}}}{1 - \frac{35}{36}}A = \left[1 - \left(\frac{35}{36}\right)^n\right] A$$

Así que se trata de obtener el más pequeño valor de n para el cual $1 - \left(\frac{35}{36}\right)^n$ sea mayor que $\frac{1}{2}$, cuya solución es $n = 25$.

Las soluciones de Huygens a los problemas resueltos por Pascal y Fermat las expuso en su libro *De ratiociniis in ludo aleae* (Del razonamiento en los juegos de azar), publicado en el año 1657. En ese libro encontramos la primera teorización del Cálculo de Probabilidades.

Huygens tomó como punto de partida la siguiente hipótesis, la cual es básicamente una definición:

En un juego, la posibilidad que se tiene de ganar alguna cosa tiene un valor tal que, si se posee ese valor, se puede uno procurar la misma posibilidad en un juego equitativo.

Por un juego equitativo, Huygens entendía un juego que no va en detrimento de ninguno de los jugadores. Incluye el caso de un juego entre un número cualquiera de jugadores en el cual, o bien todos los jugadores tienen la misma posibilidad de ganar cierta cantidad, o bien cada uno de los jugadores tiene la misma posibilidad de ganar cierta cantidad que de perderla.

De su hipótesis, Huygens dedujo tres proposiciones, de las cuales, las dos primeras son un caso particular de la siguiente:

Proposición 1. *Tener iguales posibilidades de obtener a_1, a_2, a_3, ..., a_n tiene un valor de* $\frac{a_1+a_2+a_3+\cdots+a_n}{n}$.

Demostración

Llamemos P a quien tiene iguales posibilidades de obtener a_1, a_2, a_3, ..., a_n. Consideremos entonces un juego entre P y otros $n-1$ jugadores, que llamaremos Q_2, Q_3, Q_4, ..., Q_n, en el cual los n tienen las mismas posibilidades de ganar y cada uno apuesta $\frac{a_1+a_2+a_3+\cdots+a_n}{n}$. Quien gane se lleva todas las apuestas, es decir, $a_1+a_2+a_3+\cdots+a_n$. Evidentemente, éste es un juego equitativo. Si, además, para cada $k \in \{2,3,\ldots,n\}$, P acuerda con Q_k que, si alguno de los dos gana el juego, el ganador le dará al otro la cantidad a_k. Cada uno de estos acuerdos es equitativo y no va en detrimento de ninguno de los otros jugadores; así que el juego continúa siendo equitativo. Si P gana el juego, una vez que entrega lo acordado con cada uno de los otros jugadores, obtiene a_1.Si el juego lo gana Q_k, P obtiene a_k. Así que P tiene iguales posibilidades de obtener a_1, a_2, a_3, ..., a_n, en un juego equitativo.

∎

Proposición 2. *Tener r posibilidades de obtener a y s posibilidades de obtener b, las posibilidades siendo equivalentes, tiene un valor de* $\frac{ra+sb}{r+s}$.

Demostración

Llamemos P a quien tiene r posibilidades de obtener a y s posibilidades de obtener b. Consideremos entonces un juego entre P y otros $r+s-1$ jugadores, en el cual los $r+s$ tienen las mismas posibilidades de ganar y cada uno apuesta $\frac{ra+sb}{r+s}$. Quien gane se lleva todas las apuestas, es decir, $ra+sb$. Evidentemente, éste es un juego equitativo. Supongamos, además, que P acuerda con cada uno de s jugadores que, si alguno de los dos gana el juego, el ganador le dará al otro la cantidad b; llamaremos a este conjunto de s jugadores, el grupo 1. Finalmente, supongamos también que, con cada uno del resto de los jugadores, P acuerda que, si alguno de los dos gana el juego, el ganador le dará al otro la cantidad a; llamaremos a este conjunto de $r-1$ jugadores, el grupo 2. Cada uno de estos acuerdos es equitativo y no va en detrimento de ninguno de los otros jugadores; así que el juego continúa siendo equitativo. Si P gana el juego, una vez que entrega lo acordado con cada uno de los otros jugadores, obtiene a. Si el juego lo gana algún jugador del grupo 1, P obtiene b, mientras que si lo gana alguno del grupo 2, obtiene a. Como los $r+s$ jugadores tienen las mismas posibilidades de ganar, P tiene entonces r posibilidades de obtener a y s posibilidades de obtener b, en un juego equitativo.

∎

Podemos ver que la hipótesis de Huygens es básicamente la definición del concepto de esperanza de lo que se obtiene sobre lo que está en juego; es decir, la esperanza de una variable aleatoria definida como lo que obtiene el jugador.

El otro problema que el Chevalier de Méré planteó a Pascal es el siguiente:

¿Cómo deben repartirse las apuestas en un juego que se interrumpe? Por ejemplo, suponiendo que dos jugadores, A y B, apuestan 32 pesos cada uno en un juego que consiste de partidas consecutivas, en cada una de las cuales cada jugador tiene la misma posibilidad de ganarla, de tal manera que quien gane una partida acumula un punto y el juego es ganado por quien obtenga primero cuatro puntos, ¿cómo deben de repartirse las apuestas en caso de que el juego se interrumpa cuando el jugador A ha ganado dos puntos y B un punto?

Este problema fue el que más interés provocó debido a que pocos lograron encontrar la solución correcta. La solución que dieron Pascal, Fermat y Huygens consistió en encontrar la probabilidad que cada jugador tiene de ganar el juego, partiendo de la situación en la cual se interrumpe. La solución de Fermat fue básicamente la siguiente:

Al jugador A le faltan 2 partidas para ganar y al jugador B, 3 partidas, entonces el juego termina en a lo más 4 partidas adicionales. Denotando por la letra a el que A gane una partida y por la letra b el que gane B, los posibles resultados de 4 partidas son los siguientes:

$(a, a, a, a), (a, a, a, b), (a, a, b, a), (a, a, b, b), (a, b, a, a), (a, b, a, b), (a, b, b, a), (b, a, a, a),$

$(b, a, a, b), (b, a, b, a), (b, b, a, a), (b, b, b, b), (b, b, b, a), (b, b, a, b), (a, b, b, b), (b, a, b, b)$

donde, por ejemplo, (b, b, a, b) significa que A gana sólo la tercera partida y B las otras 3.

De estos 16 posibles resultados, hay 11 que hacen ganar al jugador A, a saber, (a, a, a, a), (a, a, a, b), (a, a, b, a), (a, a, b, b), (a, b, a, a), (a, b, a, b), (a, b, b, a), (b, a, a, a), (b, a, a, b), (b, a, b, a) y (b, b, a, a). Los 5 restantes, (b, b, b, b), (b, b, b, a), (b, b, a, b), (a, b, b, b) y (b, a, b, b), hacen ganar al jugador B. Por lo tanto, las apuestas se deben repartir en la proporción 11 : 5.

Podemos observar que en la solución de Fermat parece haber un problema, pues, para contar los casos en que gana cada jugador, considera que se juegan las 4 partidas y en algunos de esos casos el juego se termina antes de llegar a la cuarta. En realidad eso no representa problema ya que, por ejemplo, si A gana las dos primeras partidas (de las 4), en cuyo caso se acabaría ahí el juego, lo que hace Fermat es descomponer ese caso, es decir (a, a) en 4 casos, a saber, (a, a, a, a), (a, a, a, b), (a, a, b, a) y (a, a, b, b); el caso (a, a) tiene 1 posibilidad de 4; los cuatro casos juntos (a, a, a, a), (a, a, a, b), (a, a, b, a) y (a, b, a, a) tienen 4 posibilidades de 16; así que las proporciones son ambas iguales a $\frac{1}{4}$. Algo similar puede argumentarse en las otras situaciones. El caso (a, b, a) se descompone en (a, b, a, a) y (a, b, a, b); el caso (b, a, a) se descompone en (b, a, a, a) y (b, a, a, b); el caso (b, b, b) se descompone en (b, b, b, b) y (b, b, b, a). El objetivo de esas descomposiciones es, como lo decía Fermat, "hacer todos los azares iguales", es decir, en lenguaje moderno, todos los casos igualmente probables.

La solución de Fermat al problema de la división de apuestas deja ver que Fermat utilizaba ya lo que más tarde se llamaría la definición clásica de probabilidad, la cual se puede aplicar siempre que los posibles resultados se puedan considerar equiprobables:

$$P(A) = \frac{\#\text{ de posibles resultados que producen la ocurrencia de } A}{\#\text{ total de posibles resultados}}$$

Jacques Bernoulli

En el año 1713 se publicó un libro de Jacques Bernoulli (1654-1705), titulado *Ars Conjectandi* (El arte de conjeturar). En ese libro, Bernoulli sentó las bases para el desarrollo posterior del Cálculo de Probabilidades. Formuló métodos generales para resolver problemas de probabilidad, dando así una formulación teórica más sólida que la de Huygens. Esto además de enfrentar el problema de probar que la frecuencia relativa con la que se presenta un evento se aproxima a la probabilidad del mismo a medida que el número de observaciones se hace más grande; en esta búsqueda demostró el primero de los teoremas límite del Cálculo de Probabilidades, el cual daría la pauta para una investigación que se prolongó por más de 200 años, hasta llegar a una formulación general que culminaría hacia el año 1930.

Bernoulli comenzó su libro con un análisis de los problemas resueltos por Huygens en su obra *De ratiociniis in ludo aleae*. Para esto tomó como base la hipótesis de Huygens, a la cual consideró como el **principio fundamental del arte de conjeturar**. Con respecto al problema planteado por el Chevalier de Méré, concerniente al lanzamiento de un par de dados, Bernoulli planteó el uso de letras, en lugar de números, con el objeto de obtener una solución general de ese problema:

Consideremos el lanzamiento de uno o muchos dados, llamemos b al número de casos en los cuales, en cada lanzamiento, se obtiene éxito en lo que se propone, c al número de casos en los cuales no se obtiene éxito y $a = b + c$. Si P juega a ganar en un lanzamiento, tiene $a - c$ casos en los cuales tiene éxito, obteniendo entonces el total en juego, el cual considera igual a 1 para simplificar; en cualquiera de los otros c casos, no obtiene nada, de manera que el valor del juego para P es igual a $\frac{a-c}{a}$. Si juega a dos lanzamientos, en el primero tiene $a - c$ casos en los cuales obtiene $1 = \frac{a}{a}$ y c casos en los cuales regresa a la situación precedente; así que el valor del juego para P es igual a $\frac{(a-c)\frac{a}{a}+c\frac{a-c}{a}}{a} = \frac{a^2-c^2}{a^2}$. Si juega a tres lanzamientos, en el primero tiene $a - c$ casos en los cuales obtiene $1 = \frac{a^2}{a^2}$ y c casos en los cuales regresa a la situación precedente; así que el valor del juego para P es igual a $\frac{(a-c)\frac{a^2}{a^2}+c\frac{a^2-c^2}{a^2}}{a} = \frac{a^3-c^3}{a^3}$. En general, si juega a n lanzamientos, el valor del juego para P es igual a $\frac{a^n-c^n}{a^n}$.

Bernoulli agregó otro método para resolver este problema: Si P juega a obtener éxito en el primer lanzamiento, el valor del juego es igual a $\frac{a-c}{a} = \frac{b}{a}$. Si juega a obtener éxito en el segundo lanzamiento, fallando en el primero, tiene c casos en los cuales regresa a la situación del inicio; así que el valor del juego es igual a $\frac{c\frac{b}{a}}{a} = \frac{bc}{a^2}$. Si juega a obtener éxito en el tercer lanzamiento, fallando en los dos primeros, tiene c casos en los cuales regresa a la situación precedente; así que el valor del juego es igual a $\frac{c\frac{bc}{a^2}}{a} = \frac{bc^2}{a^3}$. En general, si juega a obtener éxito en el k-ésimo lanzamiento, fallando en los $k - 1$ primeros, el valor del juego es igual a $\frac{c\frac{bc}{a^2}}{a} = \frac{bc^{k-1}}{a^k}$. Por lo tanto, si juega a obtener éxito en por lo menos uno de n lanzamientos, el valor del juego es la suma de los n casos particulares descritos antes; es decir:

$$\frac{b}{a} + \frac{bc}{a^2} + \frac{bc^2}{a^3} + \cdots + \frac{bc^{n-1}}{a^n} = \frac{\frac{b}{a} - \frac{bc^{n-1}}{a^n}\frac{c}{a}}{1-\frac{c}{a}} = \frac{a^n-c^n}{a^n}$$

Bernoulli agregó un tercer método: El lanzamiento de un dado n veces es equivalente a lanzar una vez n dados. Consideremos entonces n dados, cada uno con a caras, de las cuales hay c en las que no se obtiene éxito. El número de casos que se pueden obtener al lanzar los n dados es igual a a^n, de los cuales hay c^n casos en que no se obtiene éxito con ninguno de ellos; así que hay $a^n - c^n$ casos en los cuales se obtiene éxito por lo menos con uno de los dados, en cuyo caso P obtiene 1, y c^n casos en los cuales obtiene 0; por lo tanto, el valor del juego es igual a $\frac{a^n-c^n}{a^n}$.

En el tercer método de Bernoulli, aunque, al igual que Huygens, utiliza esperanzas, se puede ver claramente la definición clásica de probabilidad, además de que resuelve el problema de la manera más simple, razonando sobre los casos en que no se obtiene éxito. Este hecho muestra que no siempre la solución más simple es la primera que se ocurre e incluso puede no ser evidente; ni Pascal, ni Fermat, ni Huygens encontraron esta forma simple de resolver el problema planteado. Lo inmediato o simple de una solución a un problema requiere, a veces, de ensayos de solución y de maduración de determinados conceptos.

En el segundo método vemos el uso de la propiedad de la aditividad finita de la función de probabilidad.

Huygens consideró otro problema con dados, el cual, al ser generalizado por Bernoulli, adquiriría una importancia central en el desarrollo del Cálculo de Probabilidades:

¿Cuántos dados se requieren lanzar para que sea más favorable obtener por lo menos dos seises?

Con relación a este problema, Bernoulli encontró que la probabilidad de obtener exactamente k seises en n lanzamientos de un dado es igual a $\binom{n}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{n-k}$. Este resultado es de fundamental importancia en su trabajo pues con él se puede calcular la probabilidad de obtener una frecuencia de seises igual a $\frac{k}{n}$ en n lanzamientos de un dado y de aquí encontrar una relación entre la frecuencia de ocurrencia de un evento y su probabilidad, para obtener lo que se llama el Teorema de Bernoulli.

Otro problema de gran importancia que planteó y resolvió Huygens en su libro es el siguiente:

Dos jugadores, P y Q, juegan a lanzar alternadamente un par de dados. El juego comienza lanzando P el par de dados, con la condición de que si obtiene una suma igual a 6, gana el juego; en caso contrario, el juego continúa lanzando Q el par de dados, con la condición de que si obtiene una suma igual a 7, gana el juego; en caso contrario el juego continúa lanzando P el par de dados bajo las condiciones iniciales. ¿Cuáles son las respectivas probabilidades que cada jugador tiene de ganar el juego?

La importancia de ese problema radica en que se refiere a un experimento el cual admite una infinidad de posibles resultados, rebasando el marco de la definición clásica de probabilidad. La solución de Huygens fue como sigue:

Sea x el valor del juego para Q y a el total de las apuestas. El valor del juego para P es entonces $a - x$. Sea además y el valor del juego para Q cuando sea su turno de lanzar los dados. Al iniciarse el juego, Q tiene 5 posibilidades de obtener 0 (cuando P obtiene una suma igual a 6) y 31 posibilidades de obtener y, por lo tanto, $x = \frac{31y}{36}$. Por otra parte, cada vez que Q tenga el turno para lanzar los dados, tiene 6 posibilidades de obtener a y 30 de obtener x, por lo tanto, $y = \frac{6a+30x}{36}$. Resolviendo el sistema de ecuaciones se obtiene $x = \frac{31}{61}a$, de manera que los valores del juego para P y Q, respectivamente, están en la proporción 30 : 31.

Obsérvese que la solución de Huygens se basa en que si P y Q no ganan el juego en su primera oportunidad, se vuelve a la situación inicial.

Bernoulli resolvió este problema estableciendo una progresión geométrica para la probabilidad que cada jugador tiene de ganar el juego:

En lugar de dos jugadores, supongamos que hay una infinidad, cada uno de los cuales tiene una oportunidad para ganar, lanzando consecutivamente el par de dados, con la condición de que el primer jugador de orden impar que obtenga 6 puntos, o el primer jugador de orden par que obtenga 7 puntos, gana el juego. Denotemos por b y c al número de casos favorables y desfavorables, respectivamente, a la obtención de 6 puntos al lanzar el par de dados; por e y f al número de casos favorables y desfavorables, respectivamente, a la obtención de 7 puntos y por a al total de casos. Las posibilidades que tiene cada jugador de ganar el juego están dadas por:

Jugador 1: b casos de un total de a casos.

Jugador 2: ce casos de un total de a^2 casos.

Jugador 3: cfb casos de un total de a^3 casos.

Jugador 4: $cfce$ casos de un total de a^4 casos.

Jugador 5: $cfcfb$ casos de un total de a^5 casos.

$\vdots$

Si suponemos ahora que todos los jugadores de orden impar son sustituidos por P y todos los jugadores de orden par son sustituidos por Q, las suertes de P y Q están dadas, respectivamente, por:

$$\frac{b}{a} + \frac{cfb}{a^3} + \frac{c^2f^2b}{a^5} + \frac{c^3f^3b}{a^7} + \cdots = \frac{\frac{b}{a}}{1-\frac{cf}{a^2}} = \frac{ab}{a^2-cf}$$

$$\frac{ce}{a^2}+\frac{cfce}{a^4}+\frac{c^2f^2ce}{a^6}+\frac{c^3f^3ce}{a^8}+\cdots=\frac{\frac{ce}{a^2}}{1-\frac{cf}{a^2}}=\frac{ce}{a^2-cf}$$

Por lo tanto, la suerte de P es a la de Q como ab es a ce; es decir, como (36) (5) es a (31) (6); o, de manera equivalente, como 30 es a 31.

Bernoulli estaba estableciendo entonces que la probabilidad de que un jugador gane el juego es igual a la suma de las probabilidades de que gane en cada uno de los posibles turnos que tiene, los cuales son una infinidad. En otras palabras, está implícita en el resultado la propiedad de σ- aditividad de la función de probabilidad.

El método de Bernoulli fue retomado un poco más adelante, en el año 1718, por Abraham de Moivre en su libro, sin embargo, aunque aparentemente era conocido, no se utilizó durante el resto del siglo XVIII y todo el XIX, de manera que la propiedad de σ-aditividad de la función de probabilidad quedó relegada en la sistematización de Laplace, la cual perduró hasta principios del siglo XX.

Las soluciones de Bernoulli a los problemas resueltos por Huygens representaron un avance significativo en el camino de dotar al Cálculo de Probabilidades de una teoría que permitiera ir resolviendo problemas cada vez más complejos. Sin embargo, la aportación central de Bernoulli la encontramos en la última parte de su libro, donde planteó un problema de singular importancia, el cual sería la base para el desarrollo posterior de la teoría durante un periodo de más de 200 años. Fue a partir de ese resultado que el Cálculo de Probabilidades comenzó a ganarse un lugar importante dentro de la Matemática.

Escribió Bernoulli en su libro:

"Parece que, para hacer una hipótesis correcta sobre un hecho cualquiera, sólo es necesario calcular exactamente el número de casos posibles y, entonces, determinar las veces que puede posiblemente ocurrir un caso más que otro. Pero aquí, inmediatamente, surge nuestra mayor dificultad, porque este procedimiento se puede aplicar únicamente a muy pocos fenómenos; de hecho, casi exclusivamente a los relacionados con los juegos de azar ... pero hay otro camino que nos conduce a lo que buscamos, y nos permite, por lo menos, hallar a posteriori lo que no podemos determinar a priori, o sea, averiguando a partir de los resultados observados en numerosos casos similares.

Ha de suponerse, a este respecto, que, bajo condiciones similares, la ocurrencia (o no ocurrencia) de un suceso en el futuro seguirá la misma pauta que se ha observado para sucesos iguales en el pasado ... Lo que aún tiene que ser averiguado es si, cuando se aumenta el número de observaciones, también se sigue aumentando la probabilidad de que la proporción registrada de casos favorables y desfavorables se aproxime a la verdadera relación ... Este es el problema que he decidido publicar aquí, después de haber trabajado sobre él durante veinte años."

Obsérvese que, en su razonamiento, Bernoulli supone que en los fenómenos aleatorios existe una regularidad, a saber, que la frecuencia relativa con la que se observa que ocurre un evento, se mantiene en el futuro como se observó en el pasado. A esta propiedad la llamaremos **principio de regularidad de la frecuencia relativa con la que ocurre un evento**.

El resultado al que hace referencia Bernoulli en su libro es el ahora llamado teorema de Bernoulli, el cual, utilizando terminología moderna, se puede enunciar como sigue:

Teorema de Bernoulli. Sea $\mathcal{E}$ un experimento aleatorio que admite t posibles resultados equiprobables y A un evento relativo a ese experimento, para el cual hay r resultados que favorecen su ocurrencia. Consideremos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento $\mathcal{E}$, de tal manera que cada repetición es independiente de las otras. Sea X_{nt} el número de veces que ocurre el evento A en las primeras nt repeticiones del experimento, entonces:

$$\text{lím}_{n\to\infty} P\left[\left|\tfrac{X_{nt}}{nt} - \tfrac{r}{t}\right| > \tfrac{1}{t}\right] = 0$$

Sin modificar lo esencial del razonamiento de Bernoulli para demostrar este resultado, se puede enunciar de la siguiente manera:

Sea $\mathcal{E}$ un experimento aleatorio que admite t posibles resultados equiprobables y A un evento relativo a ese experimento, para el cual hay r resultados que favorecen su ocurrencia. Consideremos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento $\mathcal{E}$, de tal manera que cada repetición es independiente de las otras. Sean X_n el número de veces que ocurre el evento A en las primeras n repeticiones del experimento y ε un número positivo arbitrario, entonces:

$$\text{lím}_{n\to\infty} P\left[\left|\tfrac{X_n}{n} - \tfrac{r}{t}\right| > \varepsilon\right] = 0$$

El resultado de Bernoulli hizo patente que en el modelo teórico que se estaba desarrollando se da efectivamente una correspondencia entre las probabilidades y las frecuencias con que se observan los posibles resultados de un suceso azaroso. Este resultado y otros del mismo tipo que le siguieron sentaron las bases teóricas para aplicar el Cálculo de Probabilidades al estudio de datos estadísticos. Muy pronto esta teoría comenzó a aplicarse al tratamiento de datos como los acumulados en tablas de mortalidad y natalidad.

Teorema de de Moivre-Laplace

La publicación del teorema de Bernoulli hizo renacer el interés por el Cálculo de Probabilidades, el cual, después de la publicación del trabajo de Christiaan Huygens, había quedado relegado, siendo visto únicamente como una curiosidad que tenía que ver exclusivamente con los juegos de azar.

En la búsqueda de mejorar el resultado de Bernoulli, Abraham de Moivre (1667-1754), demostró en 1733 un resultado que también sería de gran importancia en el desarrollo del Cálculo de Probabilidades. En ese año se publicó su artículo titulado *Approximatio ad Summam Terminorum Binomii* $(a+b)^n$ *in Seriem expansi*, en el cual expone un resultado que conduciría a lo que ahora se conoce como el Teorema del Límite Central. El artículo fue publicado en latín y circuló en forma privada. En el año 1738 ese artículo fue incluido en la segunda edición de su libro *The Doctrine of Chances* con el título *A Method of approximating the Sum of the Terms of the Binomial* $(a+b)^n$ *expanded into a Series, from whence are deduced some practical Rules to estimate the Degree of Assent which is to be given to Experiments.* Con terminología y notación moderna, el resultado de de Moivre puede enunciarse de la siguiente manera:

Teorema de de Moivre. Si d es un número real positivo del orden de $\sqrt{n}$ y, para cada $n \in \mathbb{N}$, X_n es una variable aleatoria con distribución binomial de parámetros n y $p = \frac{1}{2}$, entonces, para n grande, se tiene:

$$P\left[-d \leq X_n - \tfrac{1}{2}n \leq d\right] \approx \tfrac{4}{\sqrt{2\pi}} \int_0^{\frac{d}{\sqrt{n}}} e^{-2y^2} dy$$

Tomando $d = \frac{1}{2}x\sqrt{n}$, donde x es un número real positivo, la aproximación de de Moivre toma la siguiente forma:

$$P\left[-x \leq \tfrac{X_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \leq x\right] \approx \tfrac{4}{\sqrt{2\pi}} \int_0^{\frac{1}{2}x} e^{-2y^2} dy = \tfrac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}z^2} dz$$

Así que podemos expresar el resultado como sigue:

$$\text{lím}_{n\to\infty} P\left[-x \leq \tfrac{X_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \leq x\right] = \tfrac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}z^2} dz$$

Mencionaba de Moivre que su resultado se puede generalizar facilmente para cualquier valor de p. Esta generalización fue expuesta por Pierre Simon Laplace (1749-1827) en su libro *Théorie Analytique des Probabilités*, publicado en el año 1812. En notación moderna, el resultado de Laplace puede escribirse de la siguiente manera:

Teorema de de Moivre-Laplace. Si d es un número real positivo del orden de $\sqrt{n}$ y, para cada $n \in \mathbb{N}$, X_n es una variable aleatoria con distribución binomial de parámetros n y p, entonces, para n grande, se tiene:

$$P\left[-d \leq X_n - np \leq d\right] \approx \tfrac{2}{\sqrt{\pi}} \int_0^{\frac{d}{\sqrt{2np(1-p)}}} e^{-y^2} dy$$

Tomando $d = x\sqrt{np\,(1-p)}$, donde x es un número real positivo, la aproximación de Laplace toma la siguiente forma:

$$P\left[-x \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq x\right] \approx \frac{2}{\sqrt{\pi}} \int_0^{\frac{x}{\sqrt{2}}} e^{-y^2} dy = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}z^2} dz$$

Así que podemos expresar el resultado como sigue:

$$\text{lím}_{n\to\infty} P\left[-x \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq x\right] = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}z^2} dz$$

En su libro Laplace realizó una síntesis del estado del Cálculo de Probabilidades en su época, agregando sus aportaciones. Expuso ahí de manera explícita la definición clásica de probabilidad:

"Se ha visto en la introducción que la probabilidad de un evento es el cociente del número de casos que le son favorables entre el número de todos los casos posibles, cuando nada hace pensar que alguno de esos casos debe ocurrir en lugar de los otros, lo cual los hace, para nosotros, igualmente posibles. La justa apreciación de esos casos diversos es uno de los puntos más delicados del Análisis de los azares."

En seguida enunció y demostró lo que ahora se denomina la propiedad de la aditividad finita de la función de probabilidad:

"Si todos los casos no son igualmente posibles, se determinará sus posibilidades respectivas, y entonces la probabilidad del evento será la suma de las probabilidades de cada caso favorable."

Después consideró el caso en que se tienen varios eventos independientes y mostró que la probabilidad de ocurrencia de todos ellos juntos es igual al producto de sus probabilidades.

Finalmente, enunció y demostró lo que se conoce ahora como la regla del producto:

"Si los eventos simples están relacionados entre ellos de manera que la suposición de la ocurrencia del primero influye en la probabilidad de ocurrencia del segundo, se tendrá la probabilidad del evento compuesto, determinando primero la probabilidad del primer evento y después la probabilidad de que el segundo ocurra dado que el primer evento ha ocurrido."

A pesar de este desarrollo, el azar parecía un concepto que en algún momento perdería importancia pues, para los pensadores de la época, sólo era producto de nuestra ignorancia. Laplace mismo formuló esta idea de manera muy clara, en su libro *Ensayo filosófico sobre las probabilidades*, publicado en el año 1814. Dice ahí que:

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

El Cálculo de Probabilidades durante la segunda mitad del siglo XIX

La teoría matemática de la probabilidad continuó desarrollándose y fue surgiendo un nuevo concepto, de gran importancia, el de variable aleatoria:

Una **variable aleatoria** es una variable cuyo valor es aleatorio, depende del resultado del experimento aleatorio en consideración.

Lo que interesa calcular de una variable aleatoria X es la probabilidad con la que toma cada uno de sus posibles valores o la probabilidad de que tome valores en un determinado intervalo. A ese conjunto de probabilidades se le llama la distribución de la variable aleatoria.

Dos cantidades de interés para el estudio de una variable aleatoria son su **esperanza** y su **varianza**. La primera expresa el valor teórico del promedio de los valores que toma la variable aleatoria cuando el experimento aleatorio correspondiente se repite muchas veces. La segunda mide la dispersión de los valores que toma la variable aleatoria, es decir, mide el alejamiento de los valores de la variable aleatoria con respecto a su esperanza.

Si una variable aleatoria X toma valores únicamente en un conjunto finito o infinito numerable, su esperanza se define de la siguiente manera:

$$E\left[X\right] = \sum_x xP\left[X = x\right]$$

La varianza de X se suele denotar por $\sigma^2\left(X\right)$ y se define de la siguiente manera:

$$\sigma^2\left(X\right) = E\left[\left(X - E\left[X\right]\right)^2\right]$$

Continuando con el estudio de los teoremas de Bernoulli y de de Moivre, se obtuvieron generalizaciones de esos resultados. En particular, la "escuela rusa" hizo grandes aportes a partir de la segunda mitad del siglo XIX:

En el año 1867, Pafnuty Lvovich Chebyshev (1821-1894) demostró una forma general del teorema de Bernoulli (**Ley débil de los grandes números**):

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

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

donde μ es la esperanza común de $X_1, X_2, \ldots$.

En el año 1900, Aleksandr Mikhailovich Lyapunov (1857-1918) demostró una forma general del teorema de de Moivre (**Teorema del Límite Central**):

Si $X_1, X_2, \ldots$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (con tercer momento finito), entonces:

$$\lim_{n\to\infty} P\left[a < \frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} < b\right] = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}dy$$

donde μ y σ^2 son la esperanza y varianza común, respectivamente, de $X_1, X_2, \ldots$.

Chebyshev y Lyapunov demostraron estos resultados asumiendo que las variables aleatorias son discretas, es decir, que toman valores únicamente en un conjunto finito o infinito numerable.

Además de lo anterior, durante la segunda mitad del siglo XIX surgió la Mecánica Estadística con los trabajos de Krönig, Clausius, Maxwell y Boltzmann, donde el Cálculo de Probabilidades se constituyó como la herramienta fundamental para el estudio de sistemas con muchas partículas.

También fue en ese periodo cuando surgió la teoría de Mendel sobre la herencia y la teoría de Darwin sobre la evolución de las especies, la primera fundada en un modelo probabilístico y la segunda planteando que el surgimiento de nuevas especies se realiza al azar. Más aún, los estudios de datos crecieron a un ritmo acelerado con los trabajos de Bienaymé, Quetelet y Galton, entre otros.

De esta forma, **a finales del siglo XIX el azar y el Cálculo de Probabilidades eran ya parte inseparable del cuerpo científico de la época**.

El Cálculo de Probabilidades durante los primeros 30 años del siglo XX

A pesar del desarrollo que tenía el Cálculo de Probabilidades a finales del siglo XIX, no había una definición satisfactoria de la probabilidad. Eso es lo que afirmaba Henri Poincaré (1854-1912) en la primera frase del capítulo I de su libro de probabilidad, publicado en 1896:

"**No se puede dar una definición satisfactoria de la probabilidad**."

En su libro, enunció la definición clásica de probabilidad:

"La probabilidad de un evento es el cociente de los casos favorables a un evento y el número total de casos posibles", aclarando mediante algunos ejemplos que se debe agregar a dicha definición la condición de que todos los casos sean igualmente probables.

Comentó entonces que:

"La definición completa de la probabilidad es una especie de petición de principio: ¿cómo reconocer que todos los casos son igualmente probables? Aquí, una definición matemática no es posible; deberemos, en cada aplicación, hacer convenciones, decir que consideramos tal y tal caso como igualmente probables. Esas convenciones no son completamente arbitrarias, pero escapan al espíritu del matemático que no tendrá más que examinarlas, una vez que son admitidas. Así, todo problema de probabilidad ofrece dos periodos de estudio: el primero, metafísico por así decirlo, el cual legitima tal o cual convención; el segundo, matemático, que aplica a esas convenciones las reglas del cálculo."

En cuanto al azar, de ser pensado únicamente un producto de nuestra ignorancia, pasó a conceptualizarse como algo objetivo. En el mismo libro, Poincaré expresó claramente este cambio:

"... en la teoría cinética de los gases, se encuentran las conocidas leyes de Mariotte y de Gay-Lussac, gracias a la hipótesis de que las velocidades de las moléculas gaseosas varían irregularmente, es decir, al azar. Las leyes observables serían mucho menos simples, dirían los físicos, si las velocidades estuvieran arregladas por alguna ley elemental simple, si las moléculas estuvieran, como se dice, organizadas, si obedecieran a alguna disciplina.

Es gracias al azar, es decir, gracias a nuestra ignorancia, que podemos concluir; y entonces, si la palabra azar es simplemente un sinónimo de ignorancia, ¿qué querría decir eso? ¿Se traduciría entonces como sigue? Me pide usted que le prediga los fenómenos que van a producirse. Si, por desgracia, conociera las leyes de esos fenómenos, podría lograrlo únicamente mediante cálculos inextricables y debería renunciar a responderle; pero, como tengo la suerte de ignorarlas, le voy a responder en seguida. Y, lo más extraordinario, es que mi respuesta será correcta. Se requiere entonces que el azar sea más que el nombre que le damos a nuestra ignorancia.”

Agregó Poincaré en su libro básicamente lo que ya había formulado Laplace como las bases del Cálculo de Probabilidades.

Decía Poincaré que el Cálculo de Probabilidades tiene como base dos teoremas: el **teorema de las probabilidades totales** y el **teorema de las probabilidades compuestas**.

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

$P(A \wedge B) = P(B \mid A)\, P(A)$

Donde $A \vee B$ representa la ocurrencia de alguno de los dos eventos A y B (incluyendo la ocurrencia de ambos), $A \wedge B$ representa la ocurrencia simultánea de los eventos A y B y $P(B \mid A)$ es la probabilidad de ocurrencia del evento B dado que el evento A ocurre.

En particular, si A y B no pueden ocurrir simultáneamente, entonces:

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

De manera más general, si $A_1, A_2, \ldots, A_n$ son eventos tales que ningún par de ellos puede ocurrir simultánemente, entonces:

$$P(A_1 \vee A_2 \vee \ldots \vee A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

Como lo mencionamos antes, a esta propiedad se le conoce como la **propiedad de la aditividad finita**.

Poincaré recogió en su libro las inquietudes de su época acerca del Cálculo de Probabilidades: No estaba bien fundamentado.

Esta era una inquietud que había no únicamente en relación a la Probabilidad. En el Congreso Internacional de Matemáticas de 1900, David Hilbert (1862-1943) expresó esas inquietudes de la siguiente manera:

“Pienso que en cualquier lugar en donde se presenten ideas matemáticas, sea en Filosofía, sea en Geometría, sea en Física, se plantea el problema de la discusión de los principios fundamentales, base de esas ideas, y del establecimiento de un sistema simple y completo de axiomas.”

“Las investigaciones sobre los principios fundamentales de la geometría nos conducen a plantear este problema: Tratar con base en ese modelo las ramas de la Física donde las

Matemáticas juegan actualmente un papel preponderante; esas ramas de la ciencia son, antes que cualesquiera otras, el Cálculo de Probabilidades y la Mecánica.”

La invención de la Teoría de la Medida a principios del siglo XX vino a resolver el problema de la fundamentación del Cálculo de Probabilidades, surgiendo así un cuerpo teórico, puramente matemático, el cual constituye lo que ahora podemos llamar la Teoría de la Probabilidad.

Inmediatamente después del surgimiento de la teoría de la medida de Lebesgue, se dio una relación con el Cálculo de Probabilidades.

En 1904, Émile Borel (1871-1956) planteó que la integral clásica (de Riemann) es insuficiente para tratar algunos problemas de probabilidad :

Si se sabe que un número x está comprendido entre 0 y 1, ¿cuál es la probabilidad de que x sea un número racional?

Utilizando la integral de Riemann, el problema no tiene solución.

Utilizando la integral de Lebesgue, la respuesta es 0.

En un inicio la identificación de la probabilidad con una medida se hizo únicamente en los problemas que caían dentro de un esquema geométrico.

En el año 1909, se publicó un articulo de Borel el cual abrió una polémica acerca de las propiedades que debían pedirse a la función probabilidad en una formulación axiomática.En ese artículo, titulado *Les probabilités dénombrables et leurs applications arithmétiques*, decía Borel:

“Se distinguen generalmente, en los problemas de probabilidad, dos categorías principales, dependiendo de que el número de casos posibles sea finito o infinito: la primera categoría constituye lo que se llama las *probabilidades discontinuas*, o probabilidades en el dominio del discontinuo, mientras que la segunda categoría comprende las *probabilidades continuas o probabilidades geométricas*. Tal clasificación aparece como incompleta cuando se consideran los resultados de la Teoría de Conjuntos; entre la potencia de los conjuntos finitos y la potencia del continuo se encuentra la potencia de los conjuntos numerables; me propongo mostrar brevemente el interés respecto a las cuestiones de probabilidad en cuyo enunciado intervienen tales conjuntos; las llamaré, para abreviar, **probabilidades numerables**.”

Enunciaremos el resultado de Borel utilizando el concepto de **ensayo de Bernoulli**, el cual se define como un experimento aleatorio que admite únicamente dos posibles resultados: éxito y fracaso.

Teorema de Borel. Consideremos una sucesión infinita numerable de ensayos de Bernoulli y sea p_n la probabilidad de éxito en el ensayo n. Denotemos por A_∞ al evento:

A_∞: Se obtiene una infinidad de éxitos.

Entonces:

Si la serie $\sum_{n=1}^{\infty} p_n$ es convergente, $P(A_\infty) = 0$.

Si los ensayos de Bernoulli son independientes y la serie $\sum_{n=1}^{\infty} p_n$ es divergente, $P(A_\infty) = 1$.

En su razonamiento, Borel utilizó algunas de las propiedades que son equivalentes a la σ-aditividad; sin embargo él consideraba que la σ-aditividad no podía considerarse como una propiedad de cualquier función de probabilidad. Para fundamentar su afirmación, daba el siguiente ejemplo:

"**Supongamos, por ejemplo, que existe una manera de elegir de entre la colección infinita de números enteros, uno de ellos al azar, de manera que cada uno de ellos tenga la misma probabilidad, esta probabilidad deberá entonces ser nula, pero su suma debe ser igual a 1**."

El teorema de Borel tiene ahora una formulación más general, conocida como lema de Borel-Cantelli.

Obsérvese que el teorema de Borel rebasó el marco clásico ya que planteó el cálculo de la probabilidad de eventos cuya ocurrencia o no ocurrencia depende de los resultados de una infininidad de ensayos de Bernoulli.

El artículo de Borel causó un gran impacto en su época sobre todo por una aplicación de sus resultados para deducir una propiedad importante de los números reales.

Sea q es un número natural mayor que 1 y, dado $x \in (0,1)$, expresemos x en la base q:

$$x = \sum_{j=1}^{\infty} \frac{b_j}{q^j},$$

donde cada b_j es un entero no negativo menor que q.

Dado un número $b \in \{0, 1, 2, \ldots, q-1\}$ denotemos por $f_n(b)$ a la fracción que resulta de dividir entre n el número de veces que aparece b en los primeros n términos del desarrollo de x en base q. Cuando $\lim_{n\to\infty} f_n(b)$ existe, llamemos a ese límite frecuencia total de b en x.

Se dice que $x = \sum_{j=1}^{\infty} \frac{b_j}{q^j}$ es normal con respecto a la base q si dado cualquier número $b \in \{0, \ldots, q-1\}$, la frecuencia total de b en x existe y su valor es igual a $\frac{1}{q}$.

Se dice que $x \in (0,1)$ es absolutamente normal si es normal con respecto a cualquier base $q \in \{2, 3, \ldots\}$.

Borel demostró entonces el siguiente resultado:

Para cada $j \in \mathbb{N}$, seleccionemos al azar un elemento del conjunto $\{0, \ldots, q-1\}$ y definamos x como la serie $\sum_{j=1}^{\infty} \frac{b_j}{q^j}$. Entonces, la probabilidad de que x sea normal con respecto a la base q es igual a 1.

El resultado de Borel puede expresarse en la forma siguiente:

Sea $\mathcal{E}$ un experimento aleatorio y A un evento relativo a ese experimento, de probabilidad igual a p. Consideremos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento $\mathcal{E}$, de tal manera que cada repetición es independiente de las otras. Sea X_n el número de veces que ocurre el evento A en las primeras n repeticiones del experimento, entonces $P\left[\lim_{n\to\infty} \frac{X_n}{n} = p\right] = 1$.

La forma general de este resultado se conoce como **Ley Fuerte de los Grandes Números**.

Más tarde, Hausdorff formuló y demostró el resultado de Borel,acerca de los números normales, utilizando La Teoría de la Medida:

Sea q es un número natural mayor que 1, entonces la medida del conjunto de puntos en el intervalo $(0, 1)$ que son normales con respecto a la base q, es igual a 1.

Como corolario se tiene que la medida del conjunto de puntos en el intervalo $(0, 1)$ que son absolutamente normales es igual a 1.

Sin embargo, hacia el año 1914 todavía no se identificaba a cualquier función de probabilidad con una medida pues ni siquiera estaba desarrollada la teoría general de la medida en espacios abstractos. En ese momento se contaba ya con la teoría de integración de Lebesgue y la correspondiente teoría de la medida en $\mathbb{R}^n$ y eran entonces éstas las únicas medidas que al normalizarlas se consideraban probabilidades. Esto es lo que hizo Felix Hausdorff en su libro, publicado en 1914. Ahí consideró que si A y B son dos conjuntos medibles de medida finita y $A \subset B$, entonces la medida de A dividida entre la medida de B puede considerarse como la probabilidad de que un punto que se selecciona en el conjunto B pertenezca al conjunto A. También en ese libro Hausdorff demostró el teorema de Borel sobre los números normales dentro del marco de la teoría de la medida.

En el libro de Hausdorff de 1914 se considera a la probabilidad como un ejemplo y una aplicación de la teoría de la medida. Hausdorff no identificaba a una probabilidad con una medida, pero mostró que una medida normalizada tiene todas las propiedades de una probabilidad.

El libro de Hausdorff fue durante mucho tiempo la referencia estándar para la teoría de conjuntos; entonces la conexión entre la probabilidad y la teoría de la medida puede considerarse como bien establecida en la literatura matemática desde 1914.

Por otra parte, en 1913, Johann Radon había ya desarrollado una teoría general de la medida en $\mathbb{R}^n$ y en 1915, con base en el trabajo de Radon, Maurice René Fréchet extendió la teoría de la medida a espacios abstractos, definiendo las funcionales aditivas. De esta manera, se puede decir que, en ese momento, aunque posteriormente todavía se demostrarían algunos resultados importantes, ya se contaba con lo básico de una teoría general de la medida.

Sin embargo, hacia 1915, aunque Frechet ya había desarrollado una teoría de la medida en espacios abstractos, no podía hacerse una identificación automática de una función de probabilidad con una medida mientras no se resolviera el problema de la existencia de una medida asociada a cada problema de probabilidad.

En 1914, Carathéodory dio un método para construir medidas en $\mathbb{R}^n$ vía una medida exterior y este método puede extenderse al caso de medidas en espacios abstractos. Sin embargo, la definición de medidas en espacios de dimensión infinita no es un problema que se haya resuelto inmediatamente después del trabajo de Fréchet sobre la definición general de una medida.

Fue P.J. Daniell quien entre 1918 y 1920 desarrolló una teoría de integración en espacios de dimensión infinita. Daniell no se basó para esto en el resultado de Carathéodory sino que desarrolló su propio método.

Básicamente el método de Carathéodory para definir una medida consiste en partir de una medida definida sobre un álgebra de subconjuntos de un conjunto dado Ω y en extender esta medida a una σ-álgebra que contiene a los conjuntos del álgebra de la que se partió. En cambio, el método de Daniel consiste en partir de una integral. definida para una cierta familia de funciones y en extender esta integral a una familia suficientemente grande de funciones. Los dos métodos son equivalentes en el sentido de que una vez teniendo una medida se puede definir una integral e inversamente, una vez teniendo una integral se puede definir una medida.

Por otra parte, el estudio de los teoremas límite había puesto en el centro de la atención de los probabilistas a las variables aleatorias. El estudio de las variables aleatorias condujo a Richard Edler Von Mises (1883-1953) a identificar, en el año 1919, una ley de probabilidad con la función de distribución. Esta misma identificación la hizo Paul Pierre Lévy (1886-1971) en su libro *Calcul des Probabilités*, publicado en 1925, donde, además, identificaba a una función de distribución con una medida sobre $\mathbb{R}$ y a una función de distribución conjunta con una medida sobre $\mathbb{R}^n$. De esta forma, dada una sola variable aleatoria, se puede asociar a ésta una medida sobre $\mathbb{R}$; dado un número finito de variables aleatorias, se puede asociar a esa familia una medida sobre $\mathbb{R}^n$, para alguna n.

Pero, ¿**cómo asociarle una medida a una familia infinita de variables aleatorias**?

Algunos resultados parciales consistentes en asociar una medida a una familia infinita de variables aleatorias se encuentran en los trabajos de Hugo Dyonizy Steinhaus (1887-1972) y de Norbert Wiener (1894-1964).

En 1923, Steinhaus consideró una sucesión infinita de ensayos de Bernoulli, en cada uno de los cuales la probabilidad de éxito es $\frac{1}{2}$, y las variables aleatorias, $X_1, X_2, \ldots$, son tales que:

$$X_j = \begin{cases} 1 & \text{si hay éxito en el ensayo } j \\ 0 & \text{si no lo hay} \end{cases}$$

El conjunto de posibles resultados del experimento aleatorio así definido consiste entonces del conjunto de sucesiones de 0's y 1's, el cual se puede poner en correspondencia, excepto por un conjunto numerable, con el intervalo $[0, 1]$.

Definió la axiomática para el juego de cara o cruz dándole a la función de probabilidad la propiedad de σ-aditividad.

Mostró entonces que comenzando por asignar probabilidades a eventos que dependen únicamente de un número finito de ensayos, las propiedades que dio a la función de probabilidad permiten definirla (extenderla) para todos los subconjuntos Lebesque-medibles y que la medida que se obtiene es precisamente la medida de Lebesgue.

Steinhaus consideró también el problema de la convergencia de series aleatorias de la forma $\sum_{n=1}^{\infty} \pm c_n$, donde cada c_n es un número real y el signo de c_n se elige al azar.

Su modelo nuevamente consiste en identificar una sucesión infinita de signos como un punto del intervalo $[0, 1]$ y entonces nuevamente asumiendo que la función de probabilidad es $\sigma-$aditiva, mostró que la función de probabilidad es la medida de Lebesgue sobre los conjuntos Lebesgue-medibles. Con base en esto demostró que la probabilidad de convergencia de una serie así definida necesariamente es 0 ó 1.

En 1924, Norbert. Wiener consideró también el problema de la convergencia de series aleatorias, pero su método fue distinto al de Steinhaus.

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

Sea Ω es el conjunto de todas las sucesiones posibles de signos. Si φ es una función definida sobre Ω cuyos valores dependen únicamente de los primeros n signos para alguna n, Wiener definió $I(\varphi)$ como el promedio de los 2^n valores que toma φ dependiendo de los primeros n signos de la sucesión. Demostró entonces que esa funcional así definida satisface las propiedades del teorema de extensión de Daniell, de manera que dicha funcional se puede extender de manera única al conjunto de todas las funciones medibles.

Con el mismo método, entre 1921 y 1923, Wiener construyó un modelo matemático para el movimiento browniano, para lo cual definió una medida de probabilidad $\sigma-$aditiva sobre el espacio de las funciones continuas. Es este trabajo el que marcó la pauta para poder definir una medida asociada a cualquier problema de probabilidad.

El **movimiento browniano** consiste en el movimiento de un grano de polen que se coloca sobre agua. En el año 1827, al estudiar el proceso de fertilización de las flores de varias plantas, Robert Brown observó que los granos de polen se movían.

Lo que hizo Wiener fue construir una medida de probabilidad sobre el conjunto de las posibles trayectorias que puede seguir una de esas partículas colocadas sobre un fluido.

Así que, pesar de la objeción de Borel, se volvió cada vez más frecuente asumir como válida ya sea la propiedad de σ-aditividad de la función de probabilidad o bien alguna de sus formas equivalentes.

Para el año 1925 algunos autores aceptaban ya a la σ-aditividad como una propiedad general de la función de probabilidad y entonces consideraban a la probabilidad como una medida. Esto queda claro en el libro de Paul Pierre Lévy de 1925, donde, además, se define a la probabilidad en forma axiomática.

Un año antes se publicó un artículo de Lévy titulado *Les lois de probabilité dans les ensembles abstraits*, en el cual dice:

Una ley de probabilidad será naturalmente bien definida en un conjunto abstracto E si se conoce la probabilidad de todo subconjunto de E. Esta probabilidad deberá gozar de las propiedades siguientes:

1. A dos conjuntos V_1 y V_2 sin elementos comunes y al conjunto V constituido por su unión, corresponden números α_1, α_2 y α tales que $\alpha = \alpha_1 + \alpha_2$.
2. Un enunciado análogo es verdadero si se considera una infinidad numerable de conjuntos $V_1, V_2, \ldots$, sin puntos comunes dos a dos.
3. Los valores de α son siempre positivos o nulos y al conjunto E completo corresponde un valor igual a la unidad.

Decía Lévy que, utilizando el lenguaje del Cálculo Funcional, α es una funcional aditiva en el sentido de Fréchet (es decir, una medida).

Agregaba después que en la práctica se considera una ley de probabilidad como definida sin que la probabilidad α esté definida para todos los subconjuntos de E. Cita para esto el caso en que la probabilidad de un subconjunto del intervalo $[0, 1]$ está dada por su medida de Lebesgue, en cuyo caso la probabilidad únicamente está definida para los conjuntos medibles.

Como puede verse, Lévy formuló aquí la Teoría de la Probabilidad en su forma axiomática moderna. Sin embargo, aunque en ese artículo Lévy formuló un método para construir medidas en espacios de dimensión infinita, éste no era lo suficientemente general.

Sin lugar a dudas, Lévy fue el más grande probabilista del siglo XX. Publicó más de 100 artículos acerca del Cálculo de Probabilidades y los Procesos Estocásticos. Su trabajo lo sistematizó en 3 libros: 1. El ya mencionado libro de 1925, *Calcul des Probabilités*, donde, además de lo que ya dijimos antes, realizó una gran sistematización del Cálculo de Probabilidades e hizo ver toda la fuerza que tiene la función característica para tratar los teoremas límite. 2. *Théorie de l'addition des variables aleatoires*, publicado en 1937, el cual contiene una amplia discusión sobre el concepto de probabilidad y el primer estudio sistemático sobre las distribuciones infinitamente divisibles. 3. *Processus Stochastiques et Mouvement Brownien*, publicado en 1948, donde introdujo los procesos con incrementos independientes e hizo un estudio minucioso del movimiento browniano, probando resultados que sólo pudieron ser demostrados formalmente años más tarde, utilizando el Cálculo Estocástico. Se mantuvo publicando artículos acerca de estos temas hasta 1970, un año antes de su fallecimiento, a los 85 años de edad.

Paralelamente a lo anterior, el Cálculo de Probabilidades tuvo una gran difusión tanto por parte de la escuela francesa como de la escuela rusa. En particular, bajo la dirección de Émile Borel, se publicó una colección de libros acerca del Cálculo de Probabilidades y sus aplicaciones, la cual consta de 4 tomos: I. *Les principes de la théorie des probabilités.* II. *Les applications de la théorie des probabilités aux sciences mathématiques et aux sciences physiques.* III. *Les applications de la théorie des probabilités aux sciences économiques et aux sciences biologiques.* IV. *Applications diverses et conclusion.* Esta colección está compuesta por 19 libros, los cuales fueron publicados entre 1924 y 1939:

La axiomática

Como lo mencionamos anteriormente, Lévy, en su libro de 1925, asumía como válida la σ-aditividad para cualquier función de probabilidad. Algunos años después, se utilizaba ya para demostrar formas generales de los teoremas límite.

En el año 1930, para probar la ley fuerte de los grandes números, Andrey Nikolaevich Kolmogorov (1903-1987) utilizó la propiedad de σ-subaditividad de la función de probabilidad, la cual es equivalente a la σ-aditividad. Además, Kolmogorov utilizó el hecho de que la unión numerable de eventos de probabilidad cero tiene también probabilidad cero, la cual también es consecuencia de la σ-subaditividad.

Ley fuerte de los grandes números (Kolmogov). Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ. Entonces:

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

Sin embargo, la polémica sobre la propiedad de $\sigma-$aditividad de la función de probabilidad continuaba. Resalta en esta polémica una serie de artículos que publicaron Maurice Fréchet y Bruno de Finetti en el año 1930.

De Finetti consideraba que se llega a contradicciones cuando se admite la extensión del teorema sobre las probabilidades totales al caso de una sucesión infinita de eventos mutuamente excluyentes. Como ejemplo consideraba una variable aleatoria X la cual únicamente puede tomar valores en el conjunto infinito $\{\varepsilon_1, \varepsilon_2, \ldots\}$ de tal forma que todos ellos son igualmente probables. Los eventos $[X = \varepsilon_i]$ tienen entonces probabilidad cero, pero su unión tiene probabilidad 1.

Fréchet argumentaba que él ya había señalado, en sus cursos y en una memoria que se encontraba en prensa, que efectivamente la extensión del teorema sobre las probabilidades totales al caso de una sucesión infinita de eventos no es una consecuencia inevitable de los principios generales admitidos en las bases del Cálculo de Probabilidades. Pero agregaba que de Finetti únicamente había visto una de las dos alternativas: "si sus ejemplos tienen sentido, entonces tal extensión no es posible. pero la otra alternativa es que si tal extensión es posible entonces los ejemplos no tienen sentido." Fréchet prefería entonces asumir que los ejemplos de de Fineti no tienen sentido, en particular consideraba, con relación al mencionado ejemplo de de Fineti, que es imposible suponer que los posibles valores de X son igualmente probables. Continuaba argumentando que la misma alternativa se presenta en la teoría de la medida de Lebesgue, donde se tiene que restringir la familia de conjuntos a los cuales se les puede asignar una medida pues no todos los conjuntos resultan ser medibles. De la misma manera, en el ejemplo de de Fineti no es posible asignarle una probabilidad a los conjuntos $[X = \varepsilon_i]$ de tal manera que todas ellas sean iguales.

Finalmente, **en el año 1933, Kolmogorov publicó un artículo titulado *Foundations of the Theory of Probability* en el cual estableció la formulación de la Teoría de la Probabilidad que prevalece hasta nuestros días**. Dice ahí:

"Después de las publicaciones de las investigaciones de Lebesgue, las analogías entre medida de un conjunto y probabilidad de un evento y entre la integral de una función y la esperanza matemática de una variable aleatoria se hicieron evidentes. Pero para que la teoría de la probabilidad pudiera basarse en tales analogías era todavía necesario hacer las teorías de la medida y de la integración independientes de los elementos geométricos los cuales estaban en el trasfondo con Lebesgue. Esto ha sido hecho por Fréchet. Mientras que una concepción de la teoría de la probabilidad basada sobre el punto de vista general citado antes se ha dado durante algún tiempo entre ciertos matemáticos, estaba faltando una exposición completa de todo el sistema, libre de extrañas complicaciones."

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

Con este modelo Kolmogorov logró entonces articular los diferentes conceptos de la teoría de la probabilidad, como el de probabilidad condicional y la independencia de eventos y de variables aleatorias. Mostró además como los resultados fundamentales de la teoría de la probabilidad se articulan en un enfoque axiomático, exponiendo, dentro de este nuevo contexto, las leyes débil y fuerte de los grandes números.

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

Teorema de extensión de Kolmogorov. Dada cualquier familia de variables aleatorias, partiendo de sus distribuciones finito dimensionales, es posible construir un espacio de probabilidad $(\Omega, \Im, P)$ de tal manera que la medida P restringida a los eventos que dependen únicamente de un número finito de las variables aleatorias dadas coincide con la determinada por la distribución finito-dimensional correspondiente.

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

Acerca de la propiedad de σ-aditividad de la función de probabilidad

El considerar a la probabilidad como una medida (σ-aditiva) constituye únicamente una elección; bien podría elegirse definir un espacio de probabilidad como una terna $(\Omega, \Im, P)$, donde $\Im$ es un álgebra de subconjuntos de Ω y P una función finitamente aditiva definida sobre $\Im$. Tendríamos así un modelo matemático similar al de la terna de Kolmogorov y la teoría se podría desarrollar con base en ese modelo. **En cualquiera de los dos casos, con una función de probabilidad σ-aditiva o con una únicamente finitamente aditiva, la teoría que se desarrolle es puramente matemática y los resultados que se obtengan son válidos, matemáticamente, dentro de esa teoría. En los dos casos se trata únicamente de un modelo matemático de los fenómenos aleatorios. El fenómeno aleatorio en sí mismo no es matemático, ni contiene matemática alguna. El modelo matemático es una abstracción producto del pensamiento humano; es parte del simbolismo que el ser humano ha creado para tratar de entender los fenómenos naturales**.

El preguntarse si una función de probabilidad es o no σ-aditiva no tiene sentido, o si se quiere, es una trivialidad. Si en el modelo que estamos utilizando tomamos a la probabilidad como σ-aditiva, entonces lo es; de otra forma, no lo es.

Elegir la σ-aditividad como propiedad de cualquier función de probabilidad nos permite definir, de manera única, la probabilidad de cada evento, a costa de que la familia de eventos tal vez no esté formada por todos los subconjuntos del espacio muestral Ω, mientras que eligiendo únicamente la aditividad finita, podemos definir la probabilidad de cualquier subconjunto del espacio muestral, pero de diferentes maneras.

No está de más remarcar que **lo que hizo Kolmogorov no fue demostrar que toda función de probabilidad es σ-aditiva**. Recapitulemos el proceso:

Primero, a una variable aleatoria se le asocia (de hecho se le identifica) con una función (llamada función de distribución) no decreciente, la cual resulta ser continua por la derecha, gracias a que se asume que la función de probabilidad es σ-aditiva; dicho de otra forma, la σ-aditividad es parte de los axiomas, de manera se asume a priori como válida. De ahí, como decíamos, la función de distribución resulta ser continua por la derecha. De ahí que podamos definir una medida (σ-adtiva) a partir de esa función de distribución y, de hecho, se identifica a la función de distribución con la medida que genera. De esta forma, asociada con una variable aleatoria real X se construye una medida sobre los subconjuntos borelianos de $\mathbb{R}$, la cual representa a la variable aleatoria. Pero, reiteramos, esto puede hacerse gracias a que de inicio se asume como válida la σ-aditividad de la función de probabilidad.

Después se trata el caso de un número finito de variables aleatorias, $X_1, X_2, \ldots, X_n$, a las cuales se les asocia una función de distribución conjunta, a partir de la cual se puede definir una medida sobre los subconjuntos borelianos de $\mathbb{R}^n$. Nuevamente esto es posible ya que se asume que la función de probabilidad es σ-aditiva.

El siguiente paso consiste en considerar una infinidad (numerable o no numerable) de variables aleatorias y Kolmogorov demostró que, partiendo de que cada subconjunto finito, de esa infinidad de variables aleatorias, tiene asociada una función de distribución y, por lo tanto, una medida sobre los subconjuntos borelianos de $\mathbb{R}^n$, para alguna $n \in \mathbb{N}$, es posible definir una medida sobre algún espacio medible, la cual representa a la infinidad de variables aleatorias.

Lo que dice entonces el teorema de Kolmogorov es que la propiedad de σ-aditividad es consistente en el sentido de que si se asume como valida para el caso finito, entonces la σ-aditividad se puede extender al caso infinito.