

Aleksandr Liapunov
 George Buffon Emile Borel
 THOMAS BAYES
 Pierre de Fermat
 Gerolamo Cardano
 Joseph Doob
 Carl Friedrich Gauss
 Boris Gnedenko Pafnuty Chebyshev
 William Feller
 Jarl Lindeberg
 Jarl Lindeberg Paul Halmos
 Paul Lévy Galileo Galilei
 Andrei Nikolaevich Kolmogorov
 George Buffon
 Blaise Pascal
 Eugene Dynkin
 Andrei Markov THOMAS BAYES
 Norbert Wiener
 Aleksandr Liapunov Boris Gnedenko
 Eugene Dynkin Aleksandr Khinchine
 Siméon Denis Poisson
 Andrei Nikolaevich Kolmogorov
 Carl Friedrich Gauss
 Norbert Wiener
 Aleksandr Liapunov
 Eugene Dynkin Aleksandr Khinchine
 Pierre-Simon (marqués de) Laplace
 Pierre-Simon (marqués de) Laplace
 Andrei Markov George Buffon
 Gerolamo Cardano
 Abraham de Moivre
 Emile Borel
 Paul Halmos
 Blaise Pascal
 William Feller
 Gottfried Wilhem Leibniz
 Galileo Galilei
 Paul Halmos Kiyoshi Itô
 THOMAS BAYES
 Siméon Denis Poisson
 Emile Borel
 Kiyoshi Itô
 Joseph Doob
 Pierre-Simon (marqués de) Laplace
 Pafnuty Chebyshev
 Pierre de Fermat
 Boris Gnedenko

PROBABILIDAD

Liliana Blanco Castañeda



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE BOGOTÁ
FACULTAD DE CIENCIAS

2a. edición

Probabilidad

Probabilidad

Liliana Blanco Castañeda

DR. RER.NAT

Profesora Asociada

Universidad Nacional de Colombia

Facultad de Ciencias

Sede Bogotá

PROBABILIDAD

© Universidad Nacional de Colombia
Facultad de Ciencias

© Edición en Castellano: Liliana Blanco Castañeda
Decano: Ignacio Mantilla Prada
Vicedecano Académico: John C. Donato
Vicedecano de Investigación y Extensión: Luis F. Ospina

Segunda Edición, 2010
ISBN: 978-958-719-576-7

Catalogación en la publicación Universidad Nacional de Colombia

Blanco Castañeda, Liliana, 1960-
Probabilidad / Liliana Blanco Castañeda.
-- 2a. ed. -- Bogotá : Universidad Nacional de Colombia.
Facultad de Ciencias, 2010
xii, 432 p., il.

Incluye referencias bibliográficas

ISBN : 978-958-719-576-7

1. Probabilidades 2. Variables aleatorias
3. Procesos de Markov I. Tít.

CDD-21 519.2 / 2010

A ti, Paula

Prefacio a la segunda edición

Enseñar bien las matemáticas solo puede hacerlo aquél que las ame con pasión y las comprenda como una ciencia viva en desarrollo.

Andréi Nikoláyevich Kolmogorov

En esta segunda edición del texto de probabilidad se han hecho varias correcciones de errores detectados en la primera edición. En el segundo capítulo se incluyó el teorema de descomposición de Jordan para funciones de distribución. En el quinto capítulo se hicieron algunas modificaciones en el orden de presentación, tanto de conceptos como de resultados, con el fin de hacer la lectura más dinámica y estructurada. En el sexto capítulo se adicionaron las principales propiedades de la esperanza condicional con respecto a una sigma álgebra.

En esta segunda edición se incluyen nuevos ejercicios en todas las secciones y se adiciona un nuevo capítulo en el que se presentan los resultados básicos de las cadenas de Markov con parámetro de tiempo discreto.

Deseo agradecer de manera especial al profesor Ignacio Mantilla, quien me colaboró con la revisión y levantamiento del texto de esta segunda edición, así como a los profesores Viswanathan Arunachalam del Departamento de Matemáticas de la Universidad de los Andes y Selvamuthu Dharmaraja del Departamento de Matemáticas del Indian Institute of Technology de New Delhi (India), por sus valiosos comentarios y sugerencias que me permitieron mejorar de manera significativa la primera edición.

Prefacio a la primera edición

Este texto está diseñado para un primer curso de teoría de probabilidad en las carreras de matemáticas y estadística, así como para los estudiantes de la especialización en estadística que no cuenten con conocimientos previos en el área. La idea de escribir este texto surgió hace ya varios años, como respuesta a la necesidad de contar con un texto en español, que abarcara los principales temas que deben ser vistos por los estudiantes de las carreras antes mencionadas. Fue así como en el año 1998 publiqué una primera versión del presente texto, en la colección “Notas de clase” de la Facultad de Ciencias de la Universidad Nacional de Colombia. Luego, a partir de la experiencia otorgada por la enseñanza del curso de probabilidad durante varios semestres, en los cuales se detectaron, gracias al aporte crítico de colegas y estudiantes, varias deficiencias y carencias en esa primera versión, ésta fue corregida y aumentada hasta dar origen al presente trabajo.

El texto está dividido en ocho capítulos y cuatro apéndices. En el primer capítulo se presentan los conceptos básicos de la teoría tales como: espacios de probabilidad, eventos independientes, probabilidad condicional y regla de Bayes. En el segundo se discuten los conceptos de variable aleatoria,

distribución y función de distribución de una variable aleatoria, valor esperado, varianza, funciones generadora de momentos y característica. En el tercer y cuarto capítulo se presentan, respectivamente, las distribuciones de tipo discreto y continuo de uso más frecuente en las aplicaciones. El quinto capítulo está dedicado al estudio de los vectores aleatorios y sus distribuciones; se estudian entre otros temas, la distribución conjunta de variables aleatorias, los conceptos de independencia de variables aleatorias y de covarianza y coeficiente de correlación, se trabaja también la distribución de una función de un vector aleatorio, obteniendo como caso particular, las distribuciones de la suma, diferencia, producto y cociente de variables aleatorias. Finalmente se dedican las últimas secciones al estudio de los conceptos de valor esperado de un vector aleatorio, matriz de varianzas y covarianzas, funciones generadoras de momentos y características conjuntas y se termina con una introducción al estudio de la distribución normal multivariada. En el sexto capítulo se trabajan los conceptos de probabilidad y esperanza condicional. El estudio de los teoremas límites es el objetivo del séptimo capítulo, en él se estudian tres tipos de convergencia de sucesiones de variables aleatorias: convergencia estocástica, casi siempre y en distribución, se establecen relaciones entre ellas y se presentan las leyes débil y fuerte de los grandes números, así como el Teorema Central del Límite. En el octavo y último capítulo se hace una introducción a la modelación aleatoria. Se explica cómo obtener algoritmos que permiten calcular probabilidades, valores esperados, así como simular valores de variables aleatorias con un número contable de posibles resultados y de variables aleatorias continuas con función de distribución conocida. En el apéndice uno se hace una breve introducción a la teoría de conjuntos, haciendo especial énfasis en los conceptos de imagen directa e inversa de una función. En el segundo apéndice se presentan los conceptos básicos del análisis

combinatorio, desarrollando ejemplos ilustrativos del principio fundamental del conteo. En el tercer apéndice se presenta un resumen de los conceptos del álgebra lineal usados a lo largo del texto. Por último, el cuarto apéndice contiene las tablas de las funciones de distribución más usadas en las aplicaciones. Estas tablas fueron generadas con el programa MATLAB.

Al final de cada capítulo hay una serie de ejercicios, con ellos el lector podrá verificar su comprensión de los temas desarrollados y encontrará, en algunos casos, material adicional de estudio.

Para la comprensión del texto, el lector debe tener conocimientos sólidos del cálculo diferencial e integral en una y varias variables.

Deseo agradecer de manera especial al profesor Ignacio Mantilla, quien además de brindarme su apoyo moral, colaboró no sólo en el desarrollo del octavo capítulo, escribiendo e implementando los algoritmos que en él aparecen, sino también en la generación de las tablas que aparecen en el cuarto apéndice.

Por último deseo agradecer a la Universidad Nacional de Colombia, y en especial a su Departamento de Estadística, por brindarme el tiempo y las facilidades necesarias que hicieron posible la elaboración de este texto; a mis colegas Fabio Nieto y Humberto Mayorga de la Universidad Nacional y Fernando Ruiz del IDEAM, quienes colaboraron en la corrección del manuscrito, así como al Profesor Gustavo Rubiano, Director de Publicaciones de la Facultad de Ciencias, quien me motivó permanentemente.

Capítulo 1

Conceptos básicos

La teoría de la probabilidad ha sido relacionada desde sus comienzos con los juegos de azar, de hecho ya en los tiempos del primer emperador romano Augusto (63 A.C-14 D.C), eran comunes los juegos de azar y se hacían tablas de mortandad. Éste fue el origen de la probabilidad y la estadística. Posteriormente estas dos disciplinas se fueron separando debido a sus distintos objetivos, pero sin dejar de estar relacionadas. En el siglo XVI se sostuvieron discusiones filosóficas sobre la probabilidad, se destaca en esta época el filósofo italiano Gerolamo Cardano (1501-1576) quien fue uno de los primeros en hacer un tratamiento matemático del azar. El trabajo de Cardano sólo fue publicado en 1663 cuando hubo el intercambio de correspondencia entre el matemático francés Blaise Pascal (1623-1662) y el abogado y matemático francés Pierre de Fermat (1601-1665) acerca del famoso problema de Chévalier de Mére.

Chévalier de Mére era el apodo del noble francés Antoine Gombaud (1607-1684) quien era un jugador empedernido que acostumbraba a invitar a sus amigos a jugar a su casa distintos tipos de juegos de dados. Uno de los juegos consistía en lanzar seis veces un dado y si en por lo menos uno los dados se obtenia un " seis", Chevalier ganaba. Con dicho juego, Chevalier era sumamente exitoso, sin embargo, para muchos de sus huéspedes el juego se tornó poco atractivo pues, a la larga, ellos siempre perdían. Por esta razón Chevalier modifica la apuesta y la cambia por la siguiente: Chevalier gana si en 24 lanzamientos de dos dados aparece, por lo menos una vez, un doble seis. Dicho juego resultó muy interesante para los huéspedes de Chevalier pero no para él, quien perdió en casi todas las noches de juego. Al no poder él aclarar el resultado de este juego, decide acudir a su amigo Blaise Pascal para determinar el por qué la suerte se había puesto en su contra.

Pascal intercambió una serie de cartas con su compatriota Fermat y entre ambos solucionaron el problema pero no pudieron aclarar la aparente contradicción entre intuición y realidad.

La solución dada por Pascal y Fermat al problema de Chevalier, en terminología moderna, es la siguiente:

Primer juego:

Chevalier gana si sale por lo menos un seis en cuatro lanzamientos del dado:

$$P(\text{"ningún seis"}) = \frac{5}{6}$$

$$P(\text{"ningún seis en cuatro lanzamientos"}) = \left(\frac{5}{6}\right)^4 = 0.482$$

$$P(\text{"por lo menos un seis en cuatro lanzamientos"}) = 0.518$$

Segundo juego:

Chevalier gana si obtiene por lo menos un doble seis en veinticuatro lanzamientos del dado:

$$P(\text{"ningún doble seis"}) = \frac{35}{36}$$

$$P(\text{"ningún doble seis en veinticuatro lanzamientos"}) = \left(\frac{35}{36}\right)^{24} = 0.509$$

$$P(\text{"por lo menos un doble seis en veinticuatro lanzamientos"}) = 0.491$$

Por lo tanto, las probabilidades de ganar de Chevalier, en la primera versión del juego, eran superiores al 50 %, en tanto que en la segunda versión estaban por debajo de dicho porcentaje y en consecuencia dicha opción le ofrecía peores posibilidades de ganar.

Este no fue el único problema que trabajaron Pascal y Fermat a través de su intercambio de correspondencia. Ellos también observaron que la probabilidad de un evento imposible debía ser cero y que la de un evento seguro debía ser uno. También observaron que en experimentos, como los del lanzamiento de un dado, la suma de todos los posibles resultados debe ser uno, estableciendo de esta manera las bases para la definición de independencia de eventos. Vale la pena anotar que ni Fermat ni Pascal publicaron sus resultados y fue el matemático y físico holandés Christian Huygens (1629-1695) quien en 1657 publicó un tratado titulado "*De ratiocinnis in ludo aleae*" (*Sobre los razonamientos relativos a los juegos de dados*) el cual estaba inspirado en la correspondencia sostenida por Fermat y Pascal. Huygens extendió los resultados de Pascal para el caso de tres o más jugadores.

En la época de Galileo Galilei (1564-1642) se plantearon muchas inquietudes de tipo combinatorio. Dentro de éstas se destaca la planteada por el conde de Toscana quien deseaba saber por qué cuando se lanzan tres dados, se tiene que la suma 10, en los resultados obtenidos, aparece más frecuentemente que la suma 9, a pesar de que hay, en cada caso, seis formas posibles de obtener el correspondiente resultado. Ese problema ya había sido planteado un siglo atrás por otras personas, pero sólo Galileo pudo resolverlo correctamente. Galileo notó que no sólo debía tenerse en cuenta la suma final de los resultados sino también cuáles son las diferentes combinaciones que permiten obtener el resultado deseado.

Este tipo de reflexiones se las hizo también el matemático alemán Gottfried Wilhem Leibniz (1646-1716). Leibniz fue uno de los últimos “Científicos Universales” de su tiempo: era filósofo, matemático y se desempeñó como diplomático. Dentro de sus aportes se encuentra la invención de una máquina para calcular y el desarrollo del cálculo diferencial e integral. En el año 1666 publicó un trabajo titulado “*Dissertatio de arte combinatoria*” el cual puede considerarse como la partida de nacimiento de la combinatoria. Él consideró no sólo la combinación de números sino también la de vocablos, colores y sonidos.

Entre los siglos XVII y XVIII se hicieron importantes avances en la teoría de la probabilidad debido en parte al desarrollo del cálculo infinitesimal; de este período sobresalen, entre otros resultados, los siguientes: la ley de los grandes números debida a James Bernoulli (1654-1705), también conocido como Jacob, Jacques o Jakob Bernoulli. Esta ley es un teorema básico en la teoría moderna de probabilidad, una interpretación sencilla de ella establece que si se realiza un experimento aleatorio en el que hay sólo dos posibles resultados: éxito y fracaso, entonces la proporción de éxitos obtenidos tiende a estabilizarse alrededor de un número entre 0 y 1

(que resulta ser la probabilidad de éxito), a medida que aumenta el número de repeticiones. El otro resultado trascendental de esta época es el teorema de DeMoivre-Laplace, el cual establece que cuando n es suficientemente grande, una variable aleatoria binomial con parámetros n y p tiene aproximadamente la misma distribución que una variable aleatoria normal con media np y varianza $np(1 - p)$. Este resultado fue probado inicialmente por el matemático francés Abraham DeMoivre (1667-1754) para el caso $p = \frac{1}{2}$ en el año 1733 y luego extendido, al caso $0 < p < 1$ arbitrario por el matemático francés Pierre-Simon Laplace (1749-1827) en el año 1812.

A pesar de haberse obtenido resultados teóricos de la envergadura de los mencionados anteriormente, es necesario anotar que en esta época no había claridad en los conceptos básicos. Para entonces, ya era conocida la famosa definición de Laplace del concepto de probabilidad como el cociente de casos favorables sobre casos posibles, suponiendo que todos los resultados del experimento aleatorio subyacente son igualmente probables; pero ¿qué significaba: “igualmente probables” ? En el año 1892, el matemático y filósofo alemán Karl Stumpf (1848-1936) interpreta esta expresión afirmando que los eventos son igualmente probables, cuando no se tiene ningún conocimiento de cuál de los distintos resultados del experimento en cuestión va a ocurrir. En contraposición con este punto de vista, se tiene el del matemático alemán Johann von Kriess (1853-1928), quien consideraba que para determinar los resultados igualmente probables de un experimento, era necesario tener un conocimiento objetivo de éste. Así por ejemplo, si sólo se sabe que una urna contiene bolas negras y blancas, entonces, según Strumpf, para la primera extracción es igualmente probable sacar una bola blanca o una bola negra, en tanto que von Kriess admitiría este supuesto sólo si se conoce que el número de bolas negras en la urna es igual al número de bolas blancas. Se dice que el mismo Markov tuvo dificultades al respecto: según Krengel [Kre] en el libro de texto de Markov (1912), se encuentra el siguiente ejemplo: “supóngase que en una urna hay bolas de cuatro colores 1, 2, 3 y 4 con frecuencias desconocidas a , b , c y d , entonces la probabilidad de extraer una bola de color 1 es igual a $\frac{1}{4}$, pues los cuatro colores son igualmente probables”. Como se observa, no había claridad, aún en el caso de experimentos aleatorios con un número finito de posibles resultados, acerca de su modelación matemática.

No podemos abandonar el recuento histórico del siglo XIX sin mencionar al famoso matemático alemán Carl Friedrich Gauss (1777-

1855) a quien se le conoce como “*El príncipe de las Matemáticas*”. Gauss fue un niño prodigio que a muy temprana edad demostró su increíble capacidad matemática. A la edad de 18 años, Gauss se propuso calcular la órbita de un asteroide llamado Ceres, que había sido descubierto el primer día del siglo XIX por el astrónomo italiano Giuseppe Piazzi . La técnica utilizada por Gauss consistió en demostrar cómo las variaciones en los datos de origen experimental podían representarse mediante una curva acampanada (hoy conocida como campana de Gauss) para lo cual empleó el llamado *método de los mínimos cuadrados*, el cual fue desarrollado por él mismo en 1794. Sus cálculos fueron tan precisos que el asteroide pudo ser observado un año después en el lugar predicho por Gauss. Gauss contribuyó además significativamente en muchos otros campos, incluida la teoría de números, el análisis matemático, la geometría diferencial, la geodesia, el magnetismo y la óptica.

A comienzos del siglo XX y a pesar de haber ocupado la atención de los más famosos matemáticos, como por ejemplo Cardano, Fermat, Bernoulli, Laplace, Poisson y Gauss, la teoría de la probabilidad no era reconocida dentro del medio académico como una disciplina matemática y se discutía si más bien se trataba de una disciplina empírica. En el famoso Segundo Congreso Internacional de Matemáticas, realizado en agosto del año 1900, el matemático alemán David Hilbert (1862-1943) plantea, en su transcendental conferencia del 8 de agosto, en el sexto problema la necesidad de investigar los fundamentos de aquellas ciencias en las que las matemáticas juegan un papel importante, entre las que estaba la teoría de la probabilidad. En el año 1901 el matemático alemán Georg Bohlmann (1869-1928) formuló una primera aproximación a la axiomatización de la probabilidad [Kre]: él define la probabilidad de un evento E como un número no negativo $p(E)$ para el cual se satisface:

i) Si E es el evento seguro entonces $p(E) = 1$.

ii) Si E_1 y E_2 son dos eventos, tales que ellos ocurren simultáneamente con probabilidad cero, entonces la probabilidad de que E_1 o E_2 ocurran es igual a $p(E_1) + p(E_2)$.

En el año 1907 el italiano Ugo Broggi (1880-1965) escribió, bajo la dirección de Hilbert, su trabajo de doctorado titulado “Die Axiome der Wahrscheinlichkeitsrechnung” (*Los axiomas del cálculo de probabilidades*). La definición de evento se presenta de manera difusa y se afirma que la aditividad y la σ -aditividad son equivalentes (la demostración de este resultado falso contiene tal cantidad de errores, que es de suponer que Hilbert no leyó en detalle el trabajo). Sin embargo, puede considerarse este trabajo como el predecesor de los trabajos de Kolmogorov.

En el congreso internacional de matemáticas llevado a cabo en Roma en el año 1908, Bohlman define la independencia de eventos en la forma conocida actualmente y muestra la diferencia de este concepto con el de la independencia dos a dos. Cabe anotar que aún no se había dado una definición precisa de evento.

De acuerdo a lo relatado por Krengel [Kre], en el año 1901 el matemático sueco Wiman utiliza el concepto de medida en su definición de probabilidad geométrica. A este respecto, dice el matemático francés Emile Borel (1871-1956) en el año 1905 lo siguiente: “*Cuando se usa la convención: la probabilidad de un conjunto es proporcional a su longitud, área o volumen, entonces se debe ser explícito y aclarar que esto es sólo una convención más y no una definición de probabilidad*”.

Gracias a los trabajos de los matemáticos franceses Fréchet (1878-1973) y Caratheodory (1873-1950), quienes “liberaron” la teoría de la medida de su interpretación geométrica, se abrió el camino para la axiomatización de la

teoría de la probabilidad tal y como se la conoce hoy en día. En el famoso libro “Grundbegriffe der Wahrscheinlichkeitsrechnung” (*Fundamentos de la teoría de probabilidad*) publicado por primera vez en el año 1933, el matemático ruso Andrei Nikolaevich Kolmogorov (1903-1987) axiomatizó la teoría de la probabilidad haciendo uso de la teoría de la medida, quedando rigurosamente definidos los conceptos de espacio de probabilidad, evento, variable aleatoria, independencia de eventos y de variables aleatorias, probabilidad condicional, entre otros. Pero en el trabajo realizado por Kolmogorov no sólo se establecieron de manera explícita los axiomas y definiciones básicas de la teoría del cálculo de probabilidades, sino que además se formularon las bases de la teoría de los procesos estocásticos, en particular se realizaron importantes contribuciones en el desarrollo de los procesos de Markov y de los procesos de ramificación. Uno de los resultados más importantes presentados por Kolmogorov es el teorema de consistencia, el cual es fundamental cuando se desea garantizar la existencia de procesos estocásticos como elementos aleatorios de espacios de dimensión infinita.

La teoría de la probabilidad no sólo es atractiva por ser una teoría matemática compleja sino por sus múltiples aplicaciones a otros campos del conocimiento científico. Su amplia gama de aplicaciones abarca tópicos en física, química, genética, ecología, comunicaciones, demografía y finanzas, entre otros muchos.

A principios del siglo XX, uno de los problemas científicos más importantes era comprender el llamado movimiento browniano, nombrado así en honor al botánico inglés Robert Brown (1777-1858), quien observó que las partículas de polen suspendidas en un líquido se mueven de manera incesante e irregular. Brown pensó, en un comienzo, que el movimiento se debía a la naturaleza orgánica del polen, sin embargo el mismo refuta esta

teoría al verificar, con un simple experimento, que el movimiento se presentaba también en sustancias inorgánicas.

Desde los trabajos realizados por Brown y hasta finales del siglo XIX no se sabe de otras investigaciones acerca del movimiento browniano. En 1905, el físico alemán Albert Einstein (1879-1955) publica en su artículo titulado “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen” (Acerca de la teoría de la cinética molecular del movimiento, inducido por el calor, de partículas suspendidas en un líquido) [Kah] los aspectos principales del movimiento browniano, él prueba que el movimiento de la partícula en el instante t se puede modelar por medio de la distribución normal y concluye además que el movimiento se debe a los continuos choques de la partícula con las moléculas del líquido en el cual se halla suspendida. Cabe anotar sin embargo, como el mismo Einstein lo reconoce, que él desconocía los trabajos de Brown [Nel]. La primera investigación matemática acerca del movimiento browniano es debida al matemático francés Louis Bachelier (1870-1946) quien en el año 1900 propuso en su tesis doctoral “Theorie de la spéculation” (*Teoría de la especulación*) al movimiento browniano como modelo asociado a los precios especulativos. Una de las imperfecciones del modelo propuesto por Bachelier fue que en éste los precios podían ser negativos y por esto el modelo cayó en el olvido durante largo tiempo. En el año 1960 el economista Samuelson (premio Nobel de economía del año 1970) propuso la exponencial del movimiento browniano para modelar el comportamiento de los precios que están sujetos a incertidumbre.

La estructura matemática del movimiento browniano, tal y como se la conoce hoy en día, es debida al famoso matemático norteamericano Norbert Wiener (1894-1964). Por esta razón el movimiento browniano también es llamado proceso de Wiener. Aun cuando las contribuciones de Wiener

siempre tuvieron como motivo resolver problemas prácticos, ellas trascendieron más allá de la solución de dichos problemas y dieron origen a teorías matemáticas de gran valor tales como la teoría de distribuciones, el análisis de Fourier y el análisis armónico generalizado. Los primeros artículos de Wiener sobre movimiento browniano son muy difíciles de leer y sólo el renombrado matemático francés Paul Lévy (1886-1971) pudo reconocer su importancia. Paul Lévy contribuyó, de manera notable, al desarrollo de la teoría de la probabilidad y entre sus contribuciones más importantes se destacan: el concepto de martingala, los procesos de Lévy, que incluyen, entre otros, al movimiento browniano y al de Poisson, y el teorema de continuidad de funciones características. Además, Lévy dedujo varias de las propiedades más importantes del movimiento browniano. Se dice (ver [Gor]) que ha ocurrido muchas veces, que cosas que se creían nuevos descubrimientos importantes en teoría de probabilidad ya estaban contenidos, de alguna manera, en los trabajos de Lévy.

En los años setenta del siglo pasado, se describió la fórmula de Black-Scholes y Merton, que permite la valoración de opciones de compra y venta europeas. Por este trabajo se les otorgó a Scholes y Merton el premio Nobel en economía en el año 1997 (para esta fecha Black ya había fallecido). El desarrollo dado por Black-Scholes y Merton, sin embargo, no hubiese sido posible de no haber sido por los trabajos realizados por el matemático japonés Kyosi Ito (1915-2008) publicados en los años 1940 y 1946, en los cuales presenta una serie de artículos donde introduce dos de los conceptos más importantes de la teoría moderna de probabilidad: la integral estocástica y las ecuaciones diferenciales estocásticas. Estos conceptos han sido además una herramienta esencial en muchos campos de la matemática, como por ejemplo en la teoría de las ecuaciones diferenciales parciales, así

como en las aplicaciones, no sólo a las matemáticas financieras, sino a la física teórica, la biología y la ingeniería (ver [Kor]).

Como se ve, la teoría de la probabilidad es no sólo una teoría matemática hermosa sino que ofrece una amplia gama de aplicaciones a problemas concretos de la vida real, ya que muchos fenómenos que se presentan en la naturaleza son de origen aleatorio y sólo con técnicas probabilistas pueden ser modelados.

Espero que este libro de texto motive a los estudiantes de matemáticas y estadística al estudio de la teoría de la probabilidad, no sólo por su riqueza teórica y su gran cantidad de aplicaciones, sino porque conjuga las distintas teorías matemáticas: álgebra, ecuaciones diferenciales, análisis funcional con un sólo fin: tratar de entender el mundo en el que vivimos.

1.1. Espacios de probabilidad

En esta sección se va a desarrollar el concepto de medida de probabilidad y a presentar sus propiedades básicas.

Cuando se lanza un dado corriente una vez, no se puede predecir cuál será el resultado que se va a obtener, sin embargo, se sabe que el conjunto de posibles resultados es $\{1, 2, 3, 4, 5, 6\}$. Un experimento de éste tipo es lo que se llama *experimento aleatorio*. Esto es:

Definición 1.1 (experimento aleatorio) *Un experimento se dice aleatorio si su resultado no puede ser determinado de antemano.*

Se supone que el conjunto de todos los posibles resultados de un experimento aleatorio es conocido. A este conjunto se le llama *espacio muestral*. Más precisamente:

Definición 1.2 (espacio muestral) El conjunto Ω de todos los posibles resultados de un experimento aleatorio se llama espacio muestral. Los elementos $\omega \in \Omega$ son llamados puntos muestrales.

Ejemplo 1.3 Experimento: lanzamiento de una moneda corriente. Los posibles resultados, en este caso, son “cara”= c y “sello”= s . Esto es, $\Omega = \{c, s\}$. ▲

Ejemplo 1.4 Experimento: lanzamiento de un dado corriente 3 veces consecutivas. En este caso, los posibles resultados son triplas (a, b, c) con $a, b, c \in \{1, 2, 3, 4, 5, 6\}$. Esto es,

$$\Omega = \{(a, b, c) : a, b, c \in \{1, 2, 3, 4, 5, 6\}\}. \quad \blacktriangle$$

Ejemplo 1.5 Experimento: se observa el número de veces que es necesario lanzar una moneda corriente hasta obtener por primera vez cara. En este caso:

$$\Omega = \{1, 2, \dots, \infty\},$$

obsérvese que el punto muestral “1” indica que se obtuvo cara en el primer lanzamiento, el punto muestral “4” indica que en los tres primeros lanzamientos se obtuvo “sello” y en el cuarto “cara”, el punto muestral “ ∞ ” indica que nunca se obtiene “cara”. ▲

Ejemplo 1.6 Experimento: se observa la posición de una partícula que se mueve aleatoriamente sobre el eje real. En este caso $\Omega = \mathbb{R}$. ▲

Se observa que los puntos muestrales pueden ser números, vectores, símbolos, etc. los cuales están determinados por el experimento considerado.

Definición 1.7 El espacio muestral Ω se llama discreto si es finito o numerable. Un experimento aleatorio se llama finito (discreto) si su espacio muestral es finito (discreto).

Volviendo al ejemplo 1.4, una pregunta que surge de manera natural es: ¿cuál es el “chance” que tiene un “evento”, tal como “la suma de los resultados obtenidos es mayor o igual que 2”, de ocurrir?. Esto es, ¿cuál es el “chance” de ocurrir de

$$A := \{(a, b, c) \in \Omega : a + b + c \geq 2\}?$$

Pero, ¿qué es un evento? Si se sigue la idea expuesta anteriormente, se puede pensar que un evento es simplemente un subconjunto del espacio muestral, pero entonces ¿son todos los subconjuntos de un espacio muestral eventos? La respuesta a esta pregunta es no. La clase de subconjuntos del espacio muestral para los que estará definido el “chance” que tienen de ocurrir debe tener estructura de σ -álgebra, concepto que se precisará a continuación.

Definición 1.8 (σ -álgebra) Sea $\Omega \neq \emptyset$. Una colección \mathfrak{F} de subconjuntos de Ω es una σ -álgebra sobre Ω , si y sólo si,

i) $\Omega \in \mathfrak{F}$.

ii) Si $A \in \mathfrak{F}$ entonces $A^c \in \mathfrak{F}$.

iii) Si $A_1, A_2, \dots \in \mathfrak{F}$ entonces $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{F}$.

Los elementos de \mathfrak{F} se llaman eventos.

Ejemplo 1.9 Sea $\Omega \neq \emptyset$. Entonces $\mathfrak{F}_0 = \{\emptyset, \Omega\}$ y $\wp(\Omega) := \{A : A \subseteq \Omega\}$ son σ -álgebras sobre Ω . \mathfrak{F}_0 recibe el nombre de σ -álgebra trivial y $\wp(\Omega)$ recibe el nombre de σ -álgebra total. ▲

Ejemplo 1.10 Sea $\Omega = \{1, 2, 3\}$. Entonces $\mathfrak{F} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ es una σ -álgebra sobre Ω , en tanto que $\mathfrak{R} = \{\emptyset, \{1\}, \{2\}, \{3\}, \Omega\}$ no lo es. ▲

Ejemplo 1.11 Sean Ω y $\tilde{\Omega}$ conjuntos no vacíos y $\tilde{\mathfrak{S}}$ una σ -álgebra sobre $\tilde{\Omega}$. Si $T: \Omega \rightarrow \tilde{\Omega}$ es una función, entonces la colección $T^{-1}(\tilde{\mathfrak{S}}) = \{T^{-1}(A) : A \in \tilde{\mathfrak{S}}\}$ es una σ -álgebra sobre Ω . ▲

Ejemplo 1.12 Sean $\Omega \neq \emptyset$ y $\bar{\Omega}$ un subconjunto no vacío de Ω . Si \mathfrak{S} es una σ -álgebra sobre Ω , entonces $\bar{\mathfrak{S}} = \{A \cap \bar{\Omega} : A \in \mathfrak{S}\}$ es una σ -álgebra sobre $\bar{\Omega}$, llamada la huella de \mathfrak{S} sobre $\bar{\Omega}$. ▲

Ejemplo 1.13 Sean $\Omega \neq \emptyset$ finito o numerable y \mathfrak{S} una σ -álgebra sobre Ω que contiene todos los conjuntos de la forma $\{\omega\}$ con $\omega \in \Omega$. Entonces $\mathfrak{S} = \wp(\Omega)$. ▲

Ejemplo 1.14 Si $\Omega \neq \emptyset$ y $\mathfrak{S}_1, \mathfrak{S}_2, \dots$ son σ -álgebras sobre Ω entonces $\bigcap_{i=1}^{\infty} \mathfrak{S}_i$ es una σ -álgebra sobre Ω . ▲

Ejemplo 1.15 (σ -álgebra generada) Sean $\Omega \neq \emptyset$ y L una colección de subconjuntos de Ω . Sea $M = \{\mathfrak{S} : \mathfrak{S} \text{ es una } \sigma\text{-álgebra sobre } \Omega \text{ que contiene a } L\}$. Entonces, por el ejemplo anterior, se tiene que $\sigma(L) := \bigcap_{\mathfrak{S} \in M} \mathfrak{S}$ es la menor σ -álgebra sobre Ω que contiene a L . Esta σ -álgebra se llama σ -álgebra generada por L . ▲

Ejemplo 1.16 (σ -álgebra de Borel) La menor σ -álgebra sobre \mathbb{R} que contiene todos los intervalos de la forma $(-\infty, a]$ con $a \in \mathbb{R}$ se llama σ -álgebra de Borel y se denota por B . Puesto que B es una σ -álgebra entonces los siguientes conjuntos también pertenece a B :

$$\begin{aligned}
(a, \infty) &= \mathbb{R} - (-\infty, a], \\
(a, b] &= (-\infty, b] \cap (a, \infty), \\
(-\infty, a) &= \bigcup_{n=1}^{\infty} (-\infty, a - \frac{1}{n}], \\
[a, \infty) &= \mathbb{R} - (-\infty, a), \\
(a, b) &= (-\infty, b) \cap (a, \infty), \\
[a, b] &= \mathbb{R} - ((-\infty, a) \cup (b, \infty)), \\
\{a\} &= [a, a], \\
\mathbb{N} &= \bigcup_{n=0}^{\infty} \{n\}, \\
\mathbb{Q} &= \bigcup_{r \in \mathbb{Q}} \{r\}, \\
\mathbb{Q}^c &= \mathbb{R} - \mathbb{Q}.
\end{aligned}$$

donde $a, b \in \mathbb{R}$ y $a < b$. ¿Son entonces todos los subconjuntos de \mathbb{R} conjuntos boreleanos? La respuesta a esta pregunta es no. En el libro de Royden [Roy] hay un ejemplo de un subconjunto de \mathbb{R} que no es boreleano. ▲

Ejemplo 1.17 (σ – álgebra de Borel en \mathbb{R}^n) Sean $a = (a_1, \dots, a_n)$ y $b = (b_1, \dots, b_n)$ elementos de \mathbb{R}^n con $a \leq b$, es decir, $a_i \leq b_i$ para todo $i = 1, \dots, n$. La σ – álgebra generada por todos los intervalos de la forma:

$$(a, b] := \{x = (x_1, \dots, x_n) \in \mathbb{R}^n : a_i < x_i \leq b_i, i = 1, \dots, n\}$$

se llama σ – álgebra de Borel en \mathbb{R}^n . ▲

Definición 1.18 (espacio medible) Sean $\Omega \neq \emptyset$ y \mathfrak{F} una σ – álgebra sobre Ω . La pareja (Ω, \mathfrak{F}) se llama espacio medible.

Es claro, a partir de la definición, que Ω y \emptyset pertenecen a cualquier σ – álgebra definida sobre Ω . Ω se llama evento seguro y \emptyset se llama evento imposible. Un evento de la forma $\{\omega\}$ con $\omega \in \Omega$ se llama evento elemental.

Decir que un evento A ocurre significa que el resultado obtenido, al realizar el experimento aleatorio cuyo espacio muestral es Ω , es un elemento de A . Por lo tanto si A y B son eventos entonces:

- i) $A \cup B$ es un evento que ocurre, si y sólo si, A o B o ambos ocurren.
- ii) $A \cap B$ es un evento que ocurre, si y sólo, si A y B ocurren.
- iii) A^c es un evento que ocurre, si y sólo si, A no ocurre.
- iv) $A - B$ es un evento que ocurre, si y sólo si, A ocurre pero B no.

Ejemplo 1.19 Si en el ejemplo 1.4 se consideran los eventos: $A =$ “el resultado del primer lanzamiento es un número primo” y $B =$ “ la suma de los resultados obtenidos es menor o igual a 4”, entonces:

$$A \cup B = \{(a, b, c) \in \Omega : a \in \{2, 3, 5\} \text{ o } (a + b + c) \leq 4\},$$

así por ejemplo, $(2, 1, 1)$, $(5, 3, 4)$, $(1, 1, 1)$ son elementos de $A \cup B$. Por otra parte

$$A \cap B = \{(a, b, c) : a \in \{2, 3, 5\} \text{ y } (a + b + c) \leq 4\} = \{(2, 1, 1)\}.$$

¿A qué son iguales $A - B$ y A^c ? (Ejercicio!). ▲

Definición 1.20 (eventos mutuamente excluyentes) Dos eventos A y B se dicen mutuamente excluyentes si $A \cap B = \emptyset$.

Ejemplo 1.21 Se lanza una moneda corriente una vez. Sean $A =$ “el resultado obtenido es cara” y $B =$ “el resultado obtenido es sello”. Es claro que A y B son mutuamente excluyentes. ▲

Ejemplo 1.22 Se lanza una moneda corriente tantas veces como sea necesario hasta obtener cara por primera vez y se cuenta el número de lanzamientos necesarios para ello. Si

$A :=$ “no se obtiene cara antes del tercer lanzamiento” = $\{3, 4, 5, \dots, \infty\}$ y

$B :=$ “no se obtiene cara antes del segundo lanzamiento” = $\{2,4,5,\dots, \infty\}$ entonces, A y B no son mutuamente excluyentes. ▲

1.1.1. Concepto de probabilidad

El objetivo a continuación es asignar a cada evento A un número real no negativo que indique el “chance” que tiene A de ocurrir. Supóngase que se realiza un experimento aleatorio n veces y que las condiciones en que éste se ejecuta se mantienen más o menos constantes.

Definición 1.23 (frecuencia relativa) Para cada evento A , el número $fr(A) := \frac{n(A)}{n}$ se llama frecuencia relativa de A , donde $n(A)$ indica el número de veces que ocurre el evento A .

Ejemplo 1.24 Supóngase que se lanza una moneda corriente 100 veces y que en 60 de esos lanzamientos el resultado es “cara”, entonces las frecuencias relativas de los eventos $A :=$ “el resultado es cara” y $B :=$ “el resultado es sello” son respectivamente $\frac{3}{5}$ y $\frac{2}{5}$. ▲

Ejemplo 1.25 Se lanza un dado corriente 500 veces y se obtiene 83 veces la cara con el número 3. En este caso la frecuencia relativa del evento

$$A := \text{“el resultado obtenido es 3”}$$

es igual a $\frac{83}{500}$. ▲

Desafortunadamente se tiene que para cada A fijo, $fr(A)$ no es constante pues su valor depende de n ; sin embargo se ha observado que cuando un experimento aleatorio se realiza un número suficientemente grande de veces, bajo condiciones similares, la frecuencia relativa $fr(A)$ se estabiliza alrededor de un valor específico entre 0 y 1.

Ejemplo 1.26 *Supóngase que se lanza un dado corriente un número n de veces y sea*

$A :=$ “el resultado obtenido es 3”.

La tabla siguiente resume los valores obtenidos:

n	frecuencia	frecuencia relativa
100	14	0.14
200	29	0.145
300	51	0.17
400	65	0.1625
500	83	0.166

La estabilización de la frecuencia relativa es lo que se conoce como “regularidad estadística” y es lo que le permite al hombre hacer predicciones que eliminan, aunque sea parcialmente, la incertidumbre presente en los fenómenos impredecibles.

El valor $P(A)$ alrededor del cual se estabiliza la frecuencia relativa de un evento indica el “chance” que tiene éste de ocurrir. Interesa saber cuáles propiedades tiene este número $P(A)$. En primer lugar se observa que puesto que $n(A) \geq 0$ entonces $P(A)$ debe ser mayor o igual a cero, como $n(\Omega) = n$ entonces $fr(\Omega) = 1$ y en consecuencia $P(\Omega) = 1$. Por otra parte si A y B son eventos mutuamente excluyentes entonces se satisface que $n(A \cup B) = n(A) + n(B)$ y por lo tanto $fr(A \cup B) = fr(A) + fr(B)$. Esto implica que $P(A \cup B) = P(A) + P(B)$ si $A \cap B = \emptyset$. Las anteriores observaciones inducen a dar la siguiente definición:

Definición 1.27 (espacio de probabilidad) *Sea (Ω, \mathfrak{S}) un espacio medible. Una función P definida sobre \mathfrak{S} y de valor real que satisface las siguientes condiciones:*

i) $P(A) \geq 0$ para todo $A \in \mathfrak{S}$.

ii) $P(\Omega) = 1$

iii) Si A_1, A_2, \dots son elementos de \mathfrak{F} mutuamente excluyentes, esto es

$$A_i \cap A_j = \emptyset \text{ para todo } i \neq j,$$

entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

se llama medida de probabilidad sobre (Ω, \mathfrak{F}) . La tripla $(\Omega, \mathfrak{F}, P)$ se llama espacio de probabilidad.

Ejemplo 1.28 Sean $\Omega = \{1, 2, 3\}$, $\mathfrak{F} = \{\emptyset, \Omega, \{1\}, \{2, 3\}\}$ y P la aplicación definida sobre \mathfrak{F} como sigue:

$$P(A) = \begin{cases} 1 & \text{si } 3 \in A \\ 0 & \text{si } 3 \notin A \end{cases}$$

Es fácil ver que P es una medida de probabilidad sobre (Ω, \mathfrak{F}) . ▲

Ejemplo 1.29 Sean $\Omega = \{1, 2\}$, $\mathfrak{F} = \wp(\Omega)$ y P la aplicación definida sobre \mathfrak{F} por:

$$P(A) = \begin{cases} 0 & \text{si } A = \emptyset \\ \frac{1}{3} & \text{si } A = \{1\} \\ \frac{2}{3} & \text{si } A = \{2\} \\ 1 & \text{si } A = \{1, 2\} \end{cases}$$

P es una medida de probabilidad. ▲

Ejemplo 1.30 (medida de Lebesgue) Sea $\Omega = [0, 1]$, $\overline{\mathfrak{B}}$ la huella de \mathfrak{B} en $[0, 1]$ y λ la medida de probabilidad definida sobre $\overline{\mathfrak{B}}$ que a cada intervalo finito contenido en $[0, 1]$ le asigna su longitud (una construcción precisa de λ va más