INTRODUCCIÓN A LA TEORÍA DE LA PROBABILIDAD, VOL. 1

Miguel Angel García Alvarez

Un poco de historia

Le 29 juillet 1654

Monsieur,

L'impatiente me prend aussi bien qu'a vous; et quoique je sois encore au lit, je ne puis m'empêcher de vous dire que je reçus hier au soir, de la part de M. de Carcavi, votre lettre sur les partis, que j'admire si fort, que je ne puis vous le dire. Je n'ai pas le loisir de m'etendre; mais en un mot vous avez trouvé les deux partis des dés et des parties dans la parfaite justesse; j'en suis tout satisfait; car je ne dout plus maintenant que je suis dans la verité, après la rencontre admirable où je me trouve avec vous ... j'en ai trouvé un abrégé, et proprement une autre méthode bien plus courte et plus nette, que je voudrais pouvoir vous dire ici en peu de mots; car je voudrais désormais vous ouvrir mon coeur, s'il se pouvait, tant que j'ai de joie de voir notre rencontre. Je vois bien que la verité est la même à Toulouse et à Paris.

Carta de Pascal a Fermat

El surgimiento del Cálculo de Probabilidades, como disciplina matemática independiente, tiene como base las soluciones que, durante el periodo que va del año 1654 al 1657, dieron Blaise Pascal, Pierre de Fermat ([**12**]) y Christiaan Huygens ([**14**]) a varios problemas, entre los cuales destacan los siguientes:

PROBLEMA 1. *¿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?

PROBLEMA 2. *¿Cuántas veces se necesita lanzar un par de dados para que sea más favorable obtener por lo menos un par de seises que no obtenerlo?*

PROBLEMA 3. *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?*

PROBLEMA 4. *Dos jugadores, A y B, los cuales poseen 12 fichas cada uno, juegan a lanzar sucesivamente tres dados, estableciéndose que A dará una ficha a B cada vez que se obtenga una suma igual a* 11*, mientras que B dará una ficha a A cada vez que se obtenga una suma igual a* 14*. Si el ganador del juego es el primero que llegue a poseer las* 24 *fichas, ¿cuáles son las respectivas probabilidades que cada jugador tiene de ganar el juego?*

Los problemas 1 y 2 fueron planteados a Pascal en el año 1654 por Antoine Gombaud de Méré, conocido como el chevalier de Méré, quien era aficionado a los juegos de azar y había logrado resolver el problema 2 pero no el 1. Pascal y Fermat encontraron las soluciones correctas a los dos problemas, mismas que se dieron a conocer entre ellos en una serie de cartas las cuales constituyen los únicos documentos en los cuales quedaron plasmados los métodos que utilizaron. Más tarde, Huygens, sin conocer los métodos utilizados por Pascal y Fermat, encontró también las soluciones correctas a ambos problemas y en el año 1657 publicó sus soluciones en su libro "De ratiociniis in Ludo Aleae" ([**14**]), siendo ésta la publicación que se convirtió en la base para el desarrollo posterior del Cálculo de Probabilidades.

Sin embargo, no fueron Pascal, Fermat y Huygens los primeros en resolver de manera correcta problemas de probabilidad. La historia del Cálculo de Probabilidades se remonta por lo menos al siglo X cuando se plantearon algunos problemas que más tarde fueron la base para resolver problemas de probabilidad. En particular, en esa época se planteó el problema de determinar cuántos resultados distintos pueden obtenerse al lanzar n dados. La primera solución correcta conocida de este problema se encuentra en un poema titulado "De Vetula" y escrito por Richard de Fournival (1200-1250). Ahí se afirma que 3 dados pueden caer en un total de 216 caminos.

La primera referencia conocida a una relación entre las diferentes posibilidades de ocurrencia de un evento y la frecuencia con que éste se observa, se encuentra en los comentarios a una publicación de "La Divina Comedia" que en el año 1477 hizo

Benvenuto d'Imola. Dice ahí: "Concerniente a estos lanzamientos (de dados) debe observarse que los dados son cuadrados y cualquier cara puede caer, así que un número que pueda aparecer en más caminos debe ocurrir más frecuentemente, como en el siguiente ejemplo: con tres dados, tres es el más pequeño número que puede obtenerse y sólo se obtiene con tres ases; cuatro puede obtenerse sólo en un camino, con un dos y dos ases".

En el libro titulado "Summa de Arithmetica, Geometria, Proportioniti et Proportionalità", escrito por Luca Paccioli en 1487 y publicado en 1494, se encuentra formulado un problema similar al 1: Dos personas juegan de manera que se requiere un total de 60 puntos para ganar, siendo el premio de 22 ducados. Por alguna circunstancia, cuando uno tiene acumulados 50 puntos y el otro 30, no pueden continuar el juego. ¿Qué parte del premio le corresponde a cada uno?. Paccioli consideraba, erróneamente, que la parte que corresponde a cada uno debe ser proporcional a los puntos que lleva ganados; en este caso, la repartición debería hacerse en la proporción de $5:3$, es decir, al que lleva 50 puntos le corresponderían $\frac{5}{8}$ y al otro $\frac{3}{8}$.

El primer estudio sistemático de problemas de probabilidad se debe a Girolamo Cardano, quien en el año 1526 escribió un libro titulado "Liber de Ludo Aleae", cuya primera publicación apareció en el año 1663 ([**6**]). En ese trabajo, Cardano realizó un estudio de problemas relacionados con lanzamientos de dados.

En su libro, estableció Cardano el número de posibilidades en el lanzamiento de 2 y 3 dados, obteniendo 36 y 216, respectivamente. Aunque en un lenguaje distinto al que se usó más tarde en el Cálculo de Probabilidades, Cardano planteó y resolvió, a la manera clásica, problemas de probabilidad. Un ejemplo es el siguiente:

Considerando el lanzamiento de 2 dados, estableció que por lo menos un as se obtiene de 11 maneras; lo mismo puede decirse de por lo menos un dos, y así sucesivamente. Agregaba que, sin embargo, un as o un dos no se obtiene de 22 maneras, pues hay 11 maneras en que se obtiene por lo menos un as y 9 más en que se obtiene por lo menos un dos, así que en total son 20 maneras de obtener por lo menos un as o por lo menos un dos. Continuaba diciendo que si se agrega ahora el 3, habrá 7 maneras más y así sucesivamente; en el siguiente paso habrá que sumar 5 maneras más, luego 3 y por último 1.

Decía entonces que si alguien dijera, quiero un as un dos o un tres, se sabe que hay 27 caminos favorables y como el circuito es de 36, los caminos en que no se obtiene ninguno de estos números son 9; las posibilidades son entonces de 3 a 1.

Con este razonamiento Cardano llegó de hecho a la llamada definición clásica de probabilidad estableciendo las posibilidades de obtener un determinado resultado en función del número de posibles maneras en que ese resultado puede obtenerse.

Situándonos nuevamente en la época de Pascal y Fermat, el problema 1 fue el problema que más interés provocó debido a que pocos lograron encontrar la solución correcta. La solución de Fermat a este problema es la siguiente:

Al jugador P le faltan dos partidas para ganar y al jugador Q tres partidas, entonces, a lo más en 4 partidas adicionales se acaba el juego. Denotando por la letra a el que P gane una partida y por la letra b el que gane Q, los posibles resultados de 4 partidas son los siguientes:

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

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

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

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

Los métodos seguidos por Pascal y Huygens para resolver este problema son distintos al de Fermat pero similares entre ellos. Su solución es como sigue:

Supongamos que al jugador A le falta una partida para ganar y a B dos, entonces, al jugar la siguiente partida hay dos posibilidades, la primera es que P la gane, en cuyo caso gana el juego y por lo tanto toda la apuesta, la segunda es que Q la gane, en cuyo caso P y Q quedan en igualdad de condiciones y debe entonces tocar a cada uno la mitad de las apuestas, es decir 32. Entonces en un caso a P le tocan 64 y en otro 32, así que, cualquiera que sea el caso, P tiene asegurado 32 y los otros 32 de las apuestas pueden corresponder a P o a Q con un azar igual; por lo tanto, de esos 32, la mitad debe ser para P y la otra para Q. Es decir, cuando a P le falta un punto y a Q dos, a P le corresponde $32 + 16 = 48$ y a Q 16.

Supongamos ahora que a A le falta un punto y a B tres. En esta situación, si se juega la siguiente partida, P puede ganar toda apuesta o bien 48 por el primer caso. Por lo tanto a P le corresponde $48 + \frac{1}{2}(16) = 56$ y a Q 8.

Finalmente, supongamos que a P le faltan dos puntos y a Q tres. En esa situación, si se juega la siguiente partida, P puede quedar faltándole un punto y tres a Q, en cuyo caso le corresponde 56 por el segundo caso; o bien, si Q gana esa partida, quedan en igualdad de circunstancias y toca a cada uno 32. Entonces P tiene asegurados 32 y puede ganar $56 - 32 = 24$ con un azar igual que Q; así que entonces a P le corresponde

$32 + \frac{1}{2}(24) = 44$ y a Q $8 + \frac{1}{2}(24) = 20$, es decir, la repartición de las apuestas debe ser de $11 : 5$.

Aunque los resultados de Pascal, Fermat y Huygens permitieron el establecimiento de reglas generales para resolver problemas de probabilidad y en ese sentido pueden considerarse como el origen del Cálculo de Probabilidades, la Teoría de la Probabilidad comenzó a ganarse un lugar importante dentro de la Matemática a partir del libro de Jacques Bernoulli, "Ars Conjectandi", publicado en el año 1713, ocho años después de su muerte ([**3**]).

Además de resolver con sus propios métodos los problemas ya resueltos por Pascal, Fermat y Huygens, Bernoulli se planteó un problema de singular importancia, el cual sería la base para todo el desarrollo posterior de la teoría. 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". El resultado al que hace referencia Bernoulli en su libro es el ahora llamado teorema de Bernoulli (ver sección 7.1).

Más tarde, en el año 1733 ([**10**]), siguiendo a Bernoulli, Abraham de Moivre demostraría el ahora llamado teorema de de Moivre-Laplace (ver sección 8.3). Ambos resultados constituyeron los primeros teoremas límite de la Teoría de la Probabilidad, cuyo estudio se prolongó durante un periodo de más de 200 años, sentando así las bases de la Teoría de la Probabilidad moderna.

Los teoremas límite fueron formulados y demostrados de manera general a principios del siglo XX, interviniendo en ese proceso, entre otros, Pierre Simon Laplace ([**20**], [**21**]), Siméon Denis Poisson ([**29**]), Pafnuty Lvovich Chebyshev ([**7**], [**8**], [**9**]), Andrei Andreyevich Markov ([**27**]), Aleksandr Mikhailovich Lyapunov ([**25**], [**26**]), Félix Édouard Justin Émile Borel ([**2**]), Francesco Paolo Cantelli ([**3**], [**4**], [**5**]), J. W.

Lindeberg ([**24**]), Paul Pierre Lévy ([**22**], [**23**]), Aleksandr Yakovlevich Khintchine ([**15**], [**16**]), Andrey Nikolaevich Kolmogorov ([**17**], [**18**]) y William Feller ([**11**]).

A principios del siglo XX la Teoría de la Probabilidad gozaba ya de una gran popularidad, sin embargo, sus fundamentos matemáticos no eran satisfactorios. De hecho, la probabilidad no era considerada como parte de la Matemática. Sus conceptos y métodos eran específicos para las aplicaciones y no formaban parte de una estructura abstracta general. La misma definición de probabilidad, la cual estaba basada en el concepto de equiprobabilidad, resultaba insatisfactoria pues no en todos los fenómenos aleatorios resulta evidente qué resultados pueden considerarse como equiprobables.

Una buena referencia para conocer el estado de la Teoría de la Probabilidad a principios del siglo XX es el libro de Jules Henri Poincaré ([**28**]), cuya primera frase es elocuente: "No se puede dar una definición satisfactoria de la probabilidad". Comenta más adelante 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".

El estudio de la fundamentación matemática de la Teoría de la Probabilidad se realizó en los primeros 30 años del siglo XX, hasta que, en el año 1933, A. N. Kolmogorov publicó un artículo ([**19**]) en el cual estableció la formulación de la Teoría de la Probabilidad que prevalece hasta nuestros días.

El modelo que formuló Kolmogorov es axiomático, lo cual se explica por el hecho de que, a principios del siglo XX, el método axiomático había ganado un gran prestigio, luego de las aportaciones de Nikolai Ivanovich Lobachevskii, Hermann Minkowski y otros matemáticos, las cuales mostraban que es posible definir geometrías no euclideanas mediante diferentes sistemas axiomáticos. Aportaciones como ésas, así como la búsqueda del rigor en la ciencia, habían llevado a plantear la necesidad de la axiomatización para todas las ramas de la Matemática, así como para aquellas ramas de la Física en donde las Matemáticas juegan un papel preponderante ([**13**]).

La historia de la Teoría de la Probabilidad no termina con su fundamentación matemática; ésta, que fue la conclusión de un proceso, se convirtió a su vez en punto de partida para profundizar en temas estudiados con anterioridad y para el estudio de nuevos sujetos de interés. Una vez formulado el modelo de Kolmogorov, la Teoría de la Probabilidad contó con una nueva herramienta que la haría desarrollarse mucho más: la Teoría de la Medida. En particular, la Teoría de los Procesos Estocásticos se convirtió en el centro de interés de los estudiosos de la Probabilidad, tema que hasta la fecha continúa desarrollándose.

Referencias

[1] Bernoulli, J., L'Art de Conjecturer, L.G.F. Vastel, G. Le Roy, Caen, 1801. Traducción de Ars Conjectandi, Basileae, 1713.

[2] Borel, F. E. J. E., Les probabilités dénombrables et leurs applications arithmétiques, Rendiconti del Circolo Matematico di Palermo, T. 27, p. 247-270, 1909. Reimpreso en Oeuvres de Émile Borel, Tome II, Centre National de la Recherche Scientifique, p. 1055-1079, 1972.

[3] Cantelli, F. P., Sulla legge dei grandi numeri, Mem. Acad. Lincei, Vol. 11, Série 5, p. 329-349, 1916.

[4] Cantelli, F. P., Sulla probabilità comme limite della frequenza, Rend. Acad. Lincei, Vol. 26, p. 39-45, 1917.

[5] Cantelli, F. P., Su due applicazioni di un teorema di G. Boole alla Statistica Matematica, Accademia dei Lincei Roma, Classe di Scienze Fisiche, Matematiche e Naturali, Rendiconti, 26 (5), p. 295-302, 1917.

[6] Cardano, G., Liber de ludo aleae, 1564. Publicado en Opera Imnia, Vol. 1, 1663. Traducción al inglés en The book on games on chance, Holt, Rinehart and Winston, New York, 1961.

[7] Chebyshev, P. L., Des valeurs moyennes, Matematicheskii Sbornik, 127, p. 1-9, 1867, también publicado en Liouville's Journal de Mathématiques Pures et Appliquées, 88, p.177-184, 1867.

[8] Chebyshev, P. L., Démonstration élémentaire d'une proposition générale de la théorie des probabilités.

[9] Chebyshev, P. L., Sur deux théorèmes relatifs aux probabilités.

[10] de Moivre, A., A method of aproximating 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, The doctrine of chances, Third edition, p. 243-259, A. Millar, London, 1756. Reimpreso por Chelsea, New York, 1967. Traducción (con algunas adiciones) de Approximatio ad summam terminorum binomii $(a+b)^n$ in seriem expansi, 1733.

[11] Feller, W., Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, Math. Zeitsch, 40, p. 521-559, 1935.

[12] Fermat, P. & Pascal, B., Correspondance - 1654, Oeuvres de Pascal, t. III, p. 369-430.

[13] D. Hilbert, Sur les problèmes futures des Mathématiques, Comptes Rendus du Deuxième Congrès International des mathematiciens, Paris, p. 58-114, 1900.

[14] Huygens, C., Du calcul dans les jeux de hasard, Oeuvres Complètes de Christiaan Huygens, Vol. XIV, Martinus Nijhoff, 1920. Traducción de De Ratiociniis in Aleae Ludo, 1657.

[15] Khintchine, A.Ya., Sur la loi des grands nombres, Comp. Rend. Acad. Sci., 188, p. 477-479, 1929.

[16] Khintchine, A.Ya., Sur la loi forte des grands nombres, C. R. Ac. Sc. Paris, Vol. 186, p. 285-287, 1928.

[17] Kolmogorov, A. N., Sur la loi des grands nombres, Rend. Acad. Lincei, Vol. 9, p. 470-474, 1929.

[18] Kolmogorov, A. N., Sur la loi forte des grands nombres, C. R. Ac. Sc. Paris, Vol. 191, p. 910-912, 1930.

[19] Kolmogorov, A. N., Foundations of the Theory of Probability, Chelsea, 1950. Traducción de Grundbegriffe der Wahrscheinlichkeitsrechnung, Erg Mat. 2, No. 3, 1933.
[20] Laplace, P. S., Théorie Analytique des Probabilités (1812), Livre I. Calcul des fonctions génératrices, Troisième edition, Courcier, Paris, 1820. Oeuvres complètes de Laplace, Tome septième, Gauthier-Villars, 1886.
[21] Laplace, P. S., Théorie Analytique des Probabilités (1812), Livre II. Théorie générale des probabilités, Troisième edition, Courcier, Paris, 1820. Oeuvres complètes de Laplace, Tome septième, Gauthier-Villars, 1886.
[22] Lévy, P. P., Calcul des Probabilités, Gauthier Villars, Paris, 1925.
[23] Lévy, P. P., Théorie de l'addition des variables aleatories, Gauthier Villars, Paris, 1937 (deuxième edition - 1954).
[24] Lindeberg, J. W., Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, Math. Zeitsch, t. 15, p. 211-225, 1922.
[25] Lyapunov, A. M., Sur une proposition de la Théorie des Probabilités, Izv. Akad. Nauk., Ser. 5, 13, p. 359-386, 1900.
[26] Lyapunov, A. M., Nouvelle forme du théorème sur la limite des probabilités, Notes Acad. Sci. Phys. Math. Sect., Ser. 8, 2, p. 1-24, 1901.
[27] Markov, A. A., Ischislenie Veroyatnostei (El Cálculo de Probabilidades), Moscow, 1913 (Cuarta edición, 1924).
[28] Poincaré, J. H., Calcul des Probabilités, Gauthier-Villars, París, 1896.
[29] Poisson, S. D., Recherches sur la probabilité des jugements en matière criminelle et en matière civile, Bachelier, Paris, 1837.

Parte 1

EL CÁLCULO DE PROBABILIDADES

CAPÍTULO 1

EL MODELO MATEMÁTICO

En medio de las causas variables y desconocidas que designamos con el nombre de azar y que hacen incierta e irregular la marcha de los acontecimientos, se ve surgir, a medida que ellos se multiplican, una regularidad asombrosa, que parece obedecer a un designio y que se ha considerado como una prueba de la providencia. Pero reflexionando sobre ella, se reconoce pronto que esta regularidad no es más que el desarrollo de las respectivas posibilidades de los acontecimientos simples, los cuales deben presentarse más frecuentemente cuando más probables son.

Pierre Simon Laplace

1.1. Experimentos aleatorios

El estudio de la naturaleza se realiza mediante la experimentación, es decir, la observación de sistemas que son brindados por la misma naturaleza o diseñados especialmente para el estudio de determinadas propiedades del sujeto de interés. Por ejemplo, en Física, se estudian las leyes del movimiento de los cuerpos basándose ya sea en la observación del movimiento de cuerpos que ofrece la misma naturaleza, como pueden ser los planetas de nuestro sistema solar, o bien diseñando experimentos en el laboratorio, por ejemplo, utilizando planos inclinados para estudiar el movimiento de cuerpos sobre ellos.

En general, el estudio de un determinado sistema conduce a un modelo de éste mediante el cual el estudio puede ser profundizado. El modelo, en general, es solo una aproximación del sistema real y está sujeto siempre a comprobación. Así, por ejemplo, en Física, el estudio del movimiento de los cuerpos condujo a la Mecánica Clásica, según la cual este movimiento obedece a las llamadas leyes de Newton. Éstas se aplicaron a todo sistema en donde intervienen movimientos mecánicos hasta que se descubrieron fenómenos a los cuales no se adaptaban con exactitud. La Teoría de la Relatividad mostraría que las leyes de Newton son válidas únicamente dentro de un cierto rango más allá del cual es necesario substituirlas por las leyes de la Mecánica Relativista.

En la Teoría de la Probabilidad nos planteamos el estudio de una determinada clase de experimentos mediante un modelo matemático que definiremos más adelante. Para este fin, consideraremos un experimento como cualquier proceso que conduce a un resultado específico. Así, el observar el color de los ojos de una persona, el preguntar a una persona si le gustan los chocolates, el medir el tiempo que permanece prendida una lámpara de manera ininterrumpida, el medir el tiempo que una piedra tarda en caer de un edificio, son todos ejemplos de experimentos. Como puede verse, le estamos dando al concepto de experimento un sentido muy amplio, no nos interesa el tipo de proceso involucrado, pudiendo ser éste una observación, una medición o cualquier otra cosa, lo único que requerimos es que el proceso en consideración conduzca a un resultado que puede ser especificado.

DEFINICIÓN 1.1 (**Experimento**). *Un experimento es cualquier proceso que conduce a un resultado específico.*

Para los fines del tema que abordaremos en este libro, clasificaremos a los experimentos en dos grandes categorías. Por un lado consideraremos todos aquellos experimentos en los cuales, una vez definidas todas las condiciones bajo las cuales se realizan, su resultado queda únicamente determinado. Los experimentos relativos al movimiento de un cuerpo bien determinado, sujeto a la acción de ciertas fuerzas, también bien determinadas, entran dentro de esta categoría. El movimiento del cuerpo queda perfectamente determinado conociendo la posición y velocidad del cuerpo en un momento dado, así como las fuerzas que actúan sobre él. Ese movimiento queda determinado en el sentido de que siempre que se repitan las mismas condiciones, el cuerpo seguirá la misma trayectoria.

Por otro lado, consideraremos todos aquellos experimentos en los cuales, una vez definidas las condiciones en que se realizan, su resultado no queda únicamente determinado. A un experimento de este tipo lo llamaremos aleatorio. En otras palabras, un experimento es aleatorio si, una vez definido, al considerar diferentes repeticiones de él, aunque sea potencialmente, todas bajo las condiciones establecidas, se pueden obtener resultados distintos.

DEFINICIÓN 1.2 (**Experimento aleatorio**). *Un experimento aleatorio es un experimento con la característica de que, una vez definidas todas las condiciones bajo las cuales se realiza, su resultado no queda únicamente determinado.*

Dentro de la categoría de experimentos aleatorios caben todos aquellos en los cuales la indeterminación del resultado se debe simplemente a que las condiciones en que se realizan incluyen cierta arbitrariedad que los hace impredescibles, así como aquellos que son por naturaleza no determinísticos, es decir, aquellos en los que aún cuando las condiciones en que se realizan no incluyen arbitrariedad, el resultado es impredecible por no estar únicamente determinado. No caben dentro de esta categoría de experimentos aquellos que son determinísticos pero en los cuales el experimentador no es capaz de predecirlos por ignorancia, ya sea de las condiciones en que se realizan o de las leyes que los rigen, o bien por no disponer de los medios adecuados para ello. Algunos ejemplos ilustrarán este punto:

Ejemplo 1.3. *Consideremos una línea de un boliche, boliches y una bola de boliche, todos con características físicas que podemos considerar invariables mientras experimentamos con ellos. Supongamos, además, que contamos con algún aparato que nos permite colocar los boliches en cualquier posición determinada y con otro aparato que nos permite lanzar la bola con la fuerza y dirección que deseemos. Fijemos una posición determinada de los boliches, elijamos una fuerza de lanzamiento de la bola orientada hacia los boliches y consideremos el experimento consistente en lanzar la bola con esa fuerza y dirección determinadas a priori y en observar el número de boliches que caen. Este experimento, así definido, es un ejemplo de un experimento determinístico. Incluso si el experimentador desconoce las leyes físicas involucradas en el experimento o no cuenta con los medios adecuados para realizar cálculos rápidos y entonces no puede predecir, antes de que la bola llegue a los boliches, el número de boliches que caerán, el experimento sigue siendo determinístico pues siempre que se repitan las mismas condiciones, el resultado será el mismo, es decir éste está únicamente determinado.*

Ejemplo 1.4. *Consideremos ahora el mismo dispositivo de boliches, pero definamos un nuevo experimento consistente en lanzar la bola con una fuerza y dirección arbitrarias (sobre la línea)*[1] *y en observar el número de boliches que caen. Este nuevo experimento es ahora un ejemplo de un experimento aleatorio. Incluso si el experimentador conoce perfectamente las leyes físicas involucradas en el problema y dispone de una computadora que le permite efectuar cálculos rápidos y así predecir, antes de que la bola llegue a los boliches, el número de boliches que caerán en cada lanzamiento, el experimento sigue siendo aleatorio pues la elección arbitraria de la fuerza y dirección de lanzamiento de la bola es parte del experimento y entonces el resultado de éste no está únicamente determinado.*

Ejemplo 1.5. *Supongamos que tenemos tres pelotitas pintadas, una de rojo, una de azul y otra de verde, colocadas en* 3 *cajas numeradas del* 1 *al* 3*, de manera que la bola roja se encuentre en la caja* 1*, la azul en la* 2 *y la verde en la* 3*. Consideremos entonces*

[1] cosa que se puede lograr, por ejemplo, eligiendo a una persona de manera arbitraria y pidiéndole que lanze la bola sobre la línea como mejor le parezca

el experimento consistente en pedirle a una persona, elegida arbitrariamente, que saque la pelotita colocada en la caja número 3 *y anote el color de ella como resultado del experimento. Incluso si la persona elegida desconoce la manera en que están colocadas las pelotitas en las cajas y entonces para ella el resultado del experimento es impredecible, el experimento descrito es un experimento determinístico pues estando ya colocadas las pelotitas en sus respectivas cajas, el resultado será siempre el mismo.*

EJEMPLO 1.6. *Consideremos las mismas* 3 *pelotitas del ejemplo anterior, las mismas* 3 *urnas numeradas del* 1 *al* 3 *y consideremos el experimento consistente en colocar las pelotitas en las cajas, una en cada una, de manera arbitraria, y en seleccionar después la pelotita de la urna número* 3*, anotando su color como resultado del experimento. Incluso si alguna persona fuera capaz de predecir la manera en que quedarán colocadas las pelotitas en las cajas en cada realización particular del experimento descrito, éste es un experimento aleatorio pues la colocación arbitraria de las pelotitas en las cajas es parte del experimento. La predicción del resultado en cada realización particular del experimento indicaría que cada realización particular admite un único posible resultado, pero una repetición del experimento, en las mismas condiciones (que incluyen la colocación arbitraria de las pelotitas en las cajas), puede dar un resultado distinto.*

▲

En los dos ejemplos de experimentos aleatorios dados arriba, la indeterminación del resultado de ellos se debe exclusivamente a que las condiciones en que se realizan incluyen cierta arbitrariedad y entonces, al repetirlos, podemos esperar distintos resultados. En otras palabras, si bien están definidos con base en experimentos que son por naturaleza determinísticos, se introduce la aleatoriedad al permitir que las condiciones precisas en que se realizan sean variables.

El problema de la existencia de otro tipo de experimentos aleatorios en los cuales la aleatoriedad sea intrínseca a ellos no será abordado en este libro. Los ejemplos que pueden darse para abordarlo requieren la profundización en áreas científicas como puede ser la Mecánica Cuántica. Lo único que podemos mencionar aquí es que ambos tipos de experimentos aleatorios caben en el estudio que haremos y nuestro objetivo será el establecer modelos que nos permitan estudiarlos. Como lo mencionamos con anterioridad, el único requisito que tienen que cumplir los experimentos que consideraremos es el poder especificar los resultados que se obtienen al realizarlos.

Como puede verse por los ejemplos dados arriba, en general, la definición de un experimento aleatorio involucra una familia de experimentos particulares, todos realizados bajo las mismas condiciones (las cuales puedes incluir cierta arbitrariedad). La situación es distinta únicamente cuando se trata de experimentos con una aleatoriedad intrínseca; es decir aquellos para los cuales aún cuando las condiciones bajo las cuales se realizan no incluyen arbitrariedad, el resultado no está únicamente determinado. Ahora bien, el que la definición de un experimento aleatorio involucre en general una familia de experimentos particulares no significa que sea posible realizar realmente cada experimento particular de la familia; es decir, no significa que el experimento aleatorio sea realmente repetible. El siguiente ejemplo ilustrará este punto:

Ejemplo 1.7. *Tomemos un gato para experimentar, digamos el gato de la Sra. Conchita que vive en la casa de al lado, el cual está completamente sano y no tiene ninguna herida. Consideremos entonces el experimento consistente en elegir una persona de manera arbitraria, darle una pistola cargada y pedirle que dispare sobre el gato. Como resultado del experimento, observemos si, después de realizarlo, el gato está vivo o muerto. Si quisiéramos considerar una serie de repeticiones de este experimento, en un momento dado, posiblemente después de su primera realización, este experimento ya no será repetible pues el gato ya estará herido o muerto y entonces no se pueden reproducir las condiciones originales bajo las cuales está definido el experimento. Sin embargo, el experimento que definimos es un experimento aleatorio y su definición involucra, al menos conceptualmente, una familia de experimentos particulares, a saber, diferentes personas que disparan sobre el gato de diferentes maneras particulares. El estudio de este tipo de experimentos puede ser más complejo que el de un experimento aleatorio que se pueda realizar tantas veces como queramos, sin embargo, forma parte del grupo de experimentos que nos interesa estudiar.*

▲

En general, consideraremos en este libro experimentos aleatorios del tipo de los ejemplos 1.4 y 1.6, es decir, en los cuales la aleatoriedad proviene de que las condiciones en que se realizan incluyen cierta arbitrariedad y entonces no podemos esperar el mismo resultado en cada una de sus posibles realizaciones. En ese tipo de experimentos, cada una de sus realizaciones se efectúa bajo condiciones particulares precisas y su resultado está entonces únicamente determinado; la aleatoriedad proviene de que, dentro de las condiciones generales que caracterizan al experimento aleatorio, caben diferentes condiciones precisas de experimentación. También, como nota aclaratoria, estaremos interesados fundamentalmente en el estudio de experimentos aleatorios, aunque eventualmente consideraremos experimentos determinísticos, viéndolos como casos extremos de experimentos aleatorios en los cuales sólo se admite un posible resultado. Finalmente, cabe mencionar que en todo problema de probabilidad está involucrado un experimento aleatorio, aunque no sea explícitamente.

Definición 1.8 (**Realización de un experimento aleatorio**). *A cada repetición particular de un experimento aleatorio la llamaremos una realización de éste.*

1.2. Eventos

A partir de esta sección consideraremos algunos experimentos aleatorios cuya descripción conviene señalar de una vez. En ocasiones consideraremos un experimento aleatorio consistente en lanzar cierto número de dados y anotar los números que se obtienen. En ese caso debe entenderse, a menos que se diga otra cosa, que se trata en primer lugar de dados usuales, es decir cubos con seis caras marcadas con números del 1 al 6; además se entenderá también que se trata de dados perfectamente simétricos y balanceados y que el lanzamiento de ellos se realiza también de la manera en que se hace usualmente en un juego con dados, es decir se agitan de manera arbitraria y se hacen caer sobre una mesa; finalmente se considera que el resultado que se obtiene con cada dado es el número que aparece en su cara superior. Una descripción semejante puede hacerse de un experimento aleatorio consistente en lanzar una o varias monedas y anotar el resultado que se obtiene; en ese caso, los dos posibles resultados del lanzamiento de una moneda serán llamados **cara** y **cruz**. En muchas ocasiones consideraremos un experimento aleatorio consistente en **seleccionar "al azar"** uno o varios elementos de una colección de objetos, que bien pueden ser cosas, animales, personas, etc.; en esos casos podemos hablar en general de la **elección al azar de individuos de una población** y la elección "al azar" se refiere, por el momento, a que ésta se realiza de manera arbitraria, es decir sin que haya predilección por elegir algunos de los individuos de la población. Más adelante definiremos el término "al azar" de una manera más precisa.

Hechas estas aclaraciones, consideremos el experimento aleatorio consistente en lanzar un dado 3 veces en forma consecutiva, anotando los números que se obtienen. Cada posible resultado de este experimento lo podemos representar por una terna de números; así, la terna $(2, 5, 3)$ significa que en el primer lanzamiento se obtiene 2, en el segundo 5 y en el tercero 3. Cada posible resultado de este experimento tiene determinadas propiedades, por ejemplo, el resultado $(2, 6, 1)$ tiene la propiedad de que nos da una suma igual a 9, también tiene la propiedad de que los 3 números son distintos, otra es que el número más grande es el 6, etc.; estas propiedades no son las mismas para todos los posibles resultados del experimento, aunque algunas de ellas pueden coincidir para varios resultados, por ejemplo, el resultado $(3, 2, 6)$ tiene como propiedad común, con el resultado anterior, el que su número más grande es el 6, también es común la propiedad de estar formado por 3 números distintos, en cambio la propiedad de que la suma es 9 no se presenta aquí.

Con cada posible resultado del experimento descrito podemos asociar entonces determinadas propiedades. Supongamos ahora que aún no hemos realizado el experimento y fijemos una determinada propiedad, por ejemplo, el que la suma de los 3 números que se obtienen sea 12. Antes de realizar el experimento, no sabemos si esta propiedad se presentará o no al realizarlo, únicamente realizando el experimento podemos decir si se presenta o no.

En general, diremos que una propiedad es relativa al experimento aleatorio si una vez realizado éste podemos decir si se presenta o no, es decir, si el experimento aleatorio determina la presencia o no de esa propiedad. Dado un experimento aleatorio, a cada propiedad relativa a ese experimento la llamaremos un evento relativo al experimento, o simplemente un evento, si no hay confusión sobre el experimento aleatorio al que se refiere.

DEFINICIÓN 1.9 (**Propiedad relativa a un experimento aleatorio**). *Diremos que una propiedad es relativa a un experimento aleatorio si una vez realizado éste podemos decir si se presenta o no.*

DEFINICIÓN 1.10 (**Eventos**). *Un evento es una propiedad relativa a un experimento aleatorio.*

Cuando, al realizar un experimento, la propiedad que define a un evento se presenta, diremos que el evento ocurre, en caso contrario, diremos que no ocurre. Un evento tiene entonces la característica de que una vez realizado el experimento podemos decir si ocurre o no ocurre.

DEFINICIÓN 1.11 (**Ocurrencia de un evento**). *Se dice que un evento ocurre al realizar un experimento aleatorio si la propiedad que lo caracteriza se presenta en esa realización.*

Los eventos relativos a un experimento aleatorio serán denotados por letras mayúsculas A, B, C,...; así, en el ejemplo descrito antes, podemos definir a A como el evento: 'se obtienen resultados distintos en cada uno de los 3 dados'.

A manera de ilustración, consideremos los siguientes ejemplos:

EJEMPLO 1.12. *Consideremos el experimento aleatorio consistente en lanzar* 3 *veces un dado en forma consecutiva y en anotar el número resultante en cada lanzamiento. Los siguientes son eventos relativos a este experimento:*

A*: Se obtiene una suma igual a* 9*.*

B*: Se obtiene por lo menos un* 5*.*

C*: En el segundo lanzamiento se obtiene* 3*.*

EJEMPLO 1.13. *De una urna en la cual hay* 2 *bolas rojas,* 4 *bolas blancas y* 7 *bolas negras, se extraen* 3 *bolas al azar. Los siguientes son eventos relativos a este experimento:*

A: *Se obtienen* 3 *bolas de distinto color.*

B: *Ninguna de las* 3 *bolas seleccionadas es blanca.*

Ejemplo 1.14. *Se elige al azar un matrimonio de una cierta población y se cuenta el número de hijos e hijas que ha tenido. Los siguientes son eventos relativos a este experimento:*

A: *El matrimonio seleccionado tiene* 2 *niños y una niña.*

B: *El matrimonio seleccionado tiene en total* 4 *hijos.*

C: *El matrimonio seleccionado tiene más niños que niñas.*

D: *El matrimonio seleccionado tiene únicamente un niño.*

Ejemplo 1.15. *De una determinada población humana se eligen al azar* 10 *personas y se anota su sexo, edad, peso y estatura. Los siguientes son eventos relativos a este experimento:*

A: *Todas las personas elegidas son mayores de* 21 *años.*

B: *Todas las personas elegidas menores de* 10 *años pesan menos de* 30 *kilos.*

C: *La estatura promedio de las personas elegidas es un número entre* 1.5 *m. y* 2.0 *m.*

1.3. Principio de regularidad de las frecuencias

Algo que hace interesante el estudio de los experimentos aleatorios es el hecho de que, a pesar de su aleatoriedad, presentan también una regularidad, aunque de un tipo distinto a la del ejemplo del movimiento de un cuerpo sujeto a la acción de ciertas fuerzas, en el cual la regularidad consiste en que, siempre que se repitan las condiciones establecidas para el experimento, el cuerpo seguirá la misma trayectoria.

En el caso de un experimento aleatorio, cuando, de ser posible, éste se realiza varias veces, se obtienen diferentes resultados; sin embargo, se puede observar que se manifiesta una regularidad en la frecuencia relativa de ocurrencia de cada posible resultado, manteniéndose aproximadamente constante cuando el número de realizaciones del experimento es grande. Por **frecuencia relativa** de un posible resultado o, en general, de un evento relativo a un experimento, entenderemos la fracción que resulta de dividir el número de veces que el evento ocurre en una serie de realizaciones del experimento entre el número total de veces que el experimento se realiza en esa serie. Esta frecuencia relativa se puede expresar también como un porcentaje simplemente multiplicando por 100 la fracción descrita.

Por ejemplo, consideremos el experimento aleatorio consistente en observar el sexo de un recién nacido elegido arbitrariamente en un cierto hospital. El resultado de una realización de este experimento no está únicamente determinado, pudiendo ser una de dos alternativas: hembra o varón. Sin embargo, si se mide la frecuencia relativa con que cada una de estas dos alternativas se ha presentado, se observará que la proporción de hembras y varones se mantiene aproximadamente constante.

La misma regularidad se puede observar en la frecuencia relativa de ocurrencia de cualquier evento relativo a un experimento aleatorio repetible. Así, por ejemplo, al realizar muchas veces el experimento aleatorio consistente en lanzar 3 dados y anotar los números que resultan, se podrá observar que, a medida que crece el número de repeticiones del experimento, la frecuencia relativa con que se obtiene una suma igual a 8 se mantiene aproximadamente constante.

A esta propiedad de regularidad de la frecuencia relativa con que ocurre cada uno de los posibles resultados de un experimento aleatorio repetible o, en general, cada evento relativo al experimento, la llamaremos **principio de regularidad de las frecuencias**. Este principio no es un principio absoluto que se cumple en cualquier serie de repeticiones de un experimento aleatorio; por ejemplo, supongamos que tenemos a la mano un dado perfectamente simétrico y balanceado y consideremos el experimento aleatorio consistente en lanzar ese dado de manera arbitraria sobre una mesa y en observar el número que muestra la cara superior del dado al estabilizarse sobre la mesa. Supongamos además que el experimento aleatorio descrito lo hemos ya realizado un millón de veces, observando que el número 1 resulta en 15% de los casos. Si el mismo experimento lo realizamos un millón de veces más, no necesariamente se obtendrá una frecuencia relativa de ocurrencia del número 1 cercana al 15%, en realidad cualquier frecuencia de ocurrencia es perfectamente aceptable. Sin embargo, basándonos en el principio de regularidad de las frecuencias, podemos decir que en muchas series de un millón de realizaciones del experimento aleatorio descrito, una desviación grande de la frecuencia con que ocurre el número 1 de un determinado porcentaje, ocurrirá raramente, es decir en proporción pequeña. La cadena de razonamientos en este sentido es interminable, nuevamente la última afirmación no debe tomarse como algo absoluto, siendo perfectamente aceptable que en diferentes series de un millón de realizaciones del experimento aleatorio se obtengan porcentajes de ocurrencia del número 1 completamente distintos.

Abundando un poco más en el análisis del ejemplo anterior, supongamos que en la segunda serie de un millón de realizaciones del experimento descrito, se observe que el número 1 se obtiene únicamente 10 veces, es decir que el número 1 se obtiene sólo en 0.001% de los casos en vez de un porcentaje cercano al 15%. ¿Significa eso que podemos estar seguros que el número 1 se presentará en un porcentaje muy superior al 15% en una serie posterior de realizaciones del mismo experimento, de manera que, tomando las dos series juntas, el porcentaje con que se presente el número 1 sea ya cercano al 15%? La respuesta es un no. De hecho, dos series de repeticiones del experimento descrito son totalmente independientes, es decir lo que ocurra en una de ellas no tiene ninguna influencia en lo que ocurra en la otra. Esto no contradice el que el porcentaje con que se presenta el número 1 se pueda estabilizar a la larga, pues lo que ocurre en una serie de un millón de realizaciones del experimento tiene poca influencia en el porcentaje de ocurrencia del número 1 digamos en un billón de realizaciones del mismo experimento.

Así, el principio de regularidad de las frecuencias debe interpretarse con cuidado. Expresa que la frecuencia con que se presenta cada posible resultado de un experimento aleatorio, en una serie grande de realizaciones, se mantiene aproximadamente constante, pero, una eventual desviación de esa constante, si bien es algo que ocurre raramente, es perfectamente aceptable.

1.4. El concepto de probabilidad

Consideremos un determinado experimento aleatorio y eventos A, B, C, ... relativos a ese experimento. Al considerar una realización del experimento, podemos decir si cada uno de los eventos en consideración ocurre o no, pero antes de realizar el experimento no podemos, en general, determinar si un evento ocurrirá o no al realizarlo. Sin embargo, cuando un experimento aleatorio se realiza varias veces, se puede observar que la frecuencia relativa con que ocurre cada evento, relativo a ese experimento, no es la misma para todos los eventos. La manera más simple de verificar esto es observando que en una serie de realizaciones de un experimento aleatorio hay eventos que ocurren muy raramente y otros que ocurren casi siempre. Por ejemplo, si consideramos el experimento consistente en lanzar 10 dados y lo realizamos varias veces, se podrá observar que el evento 'se obtiene un 1 en cada uno de los 10 dados' ocurre muy raramente, mientras que el evento 'la suma de los números que se obtienen en los 10 dados es mayor que 10' ocurre casi siempre. En general, la diferencia en la frecuencia relativa de ocurrencia de un par de eventos no será tan marcada como en este ejemplo, pero sí se podrá observar que, en general, son distintas para eventos distintos. Esta diferencia en la frecuencia relativa con que ocurre cada evento relativo a un experimento aleatorio puede interpretarse diciendo que, entre todos los eventos, hay unos que ocurren **"más fácilmente"** que otros. Más aún, con base en el principio de regularidad de las frecuencias, en una serie grande de realizaciones de un experimento aleatorio, la frecuencia relativa con que ocurre cada evento se mantiene aproximadamente constante. Esa constante puede interpretarse como una medida del "qué tan fácilmente" esperamos que el evento correspondiente ocurra.

El problema que se plantea en el Cálculo de Probabilidades consiste en encontrar una manera que permita medir el "qué tan fácilmente" se presentará un evento en futuras realizaciones de un experimento aleatorio. Así, se trata de asignar un número a cada evento, el cual exprese esa medida. De acuerdo con lo dicho anteriormente, esperamos que, en el caso de experimentos aleatorios repetibles, los números asignados a los diferentes eventos sean tales que, mientras más grande sea el número asignado a un evento, más grande será la frecuencia relativa con que éste ocurre en una serie grande de realizaciones del experimento.

A ese número que mide la "facilidad" con que un evento ocurre al realizar el experimento aleatorio correspondiente lo llamaremos la probabilidad del evento correspondiente.

DEFINICIÓN 1.16 (**Probabilidad de un evento**). *La probabilidad de un evento relativo a un experimento aleatorio es un número que mide la "facilidad" con que el evento ocurre al realizar el experimento.*

1.5. Espacios muestrales

En general, dado un experimento aleatorio, no es difícil precisar y denotar de alguna manera a cada uno de sus posibles resultados. Por ejemplo, vimos que en el caso del experimento aleatorio consistente en lanzar un dado 3 veces en forma consecutiva, anotando los números que se obtienen, cada posible resultado puede representarse mediante una terna de números con la convención de que, por ejemplo, la terna $(2,5,3)$ significa que en el primer lanzamiento se obtiene 2, en el segundo 5 y en el tercero 3.

A manera de notación, utilizaremos la letra ω con un subíndice para denotar a cada posible resultado de un experimento aleatorio. Así, en el caso del lanzamiento de dados descrito arriba, podríamos denotar con ω_1 al resultado $(1,1,1)$, con ω_2 al resultado $(1,1,2)$, $\ldots$.

Al conjunto formado por todos los posibles resultados de un experimento aleatorio se le llama el espacio muestral de ese experimento y se le denota con la letra Ω. Así, en el ejemplo dado arriba, el espacio muestral está dado por:

$$\Omega = \{(1,1,1),(1,1,2),...\}$$

O bien, si hemos definido a los posibles resultados usando ω's, podemos escribir:

$$\Omega = \{\omega_1, \omega_2, \ldots\}$$

Definición 1.17 (**Espacio muestral**). *El espacio muestral de un experimento aleatorio es el conjunto formado por todos sus posibles resultados.*

El espacio muestral de un experimento aleatorio puede ser un conjunto muy grande e incluso puede contener una infinidad de elementos. Un ejemplo típico de esa situación es el siguiente:

Ejemplo 1.18. *Consideremos el experimento aleatorio consistente en lanzar una moneda al aire tantas veces como sea necesario hasta obtener por primera vez cruz, anotando el resultado de cada lanzamiento. Los posibles resultados de este experimento pueden representarse por sucesiones de A's y C's, en donde A representa la obtención de cara y C la obtención de cruz en un lanzamiento. Así, la sucesión $CCCCCCA$ representa un posible resultado del experimento aleatorio descrito. Utilizando ω's, podríamos definir:*

$$\begin{array}{ccl} \omega_1 & = & C \\ \omega_2 & = & AC \\ \omega_3 & = & AAC \\ \vdots & \vdots & \vdots \end{array}$$

El número de cruces que resultan antes de la obtención de una cara puede ser cualquier número entero no negativo, es decir, en este caso el espacio muestral Ω es un conjunto infinito numerable.

▲

Como se muestra en el ejemplo siguiente, el espacio muestral de un experimento aleatorio puede incluso ser más grande que un conjunto infinito numerable.

Ejemplo 1.19. *Considérese el experimento aleatorio consistente en elegir al azar una lámpara de la producción de una cierta fábrica y en medir su tiempo de vida, es decir, el tiempo que permanece iluminando continuamente antes de fundirse. Aunque el tiempo de vida lo podríamos dar mediante un múltiplo entero de una unidad de tiempo elegida previamente, teóricamente ese tiempo de vida puede ser cualquier número real no negativo. En otras palabras, en este caso, cada posible resultado del experimento aleatorio puede representarse con un número real no negativo, que representa el tiempo de vida de la lámpara seleccionada, y $\Omega = \{x \in \mathbb{R} : x \geq 0\}$. Así, en este caso, Ω es un conjunto infinito no numerable.*

1.6. Representación de eventos

Consideremos un experimento aleatorio y un evento A relativo a ese experimento. Antes de realizar el experimento, no sabemos si el evento A ocurrirá o no. Siendo el evento A una propiedad relativa al experimento, solo podemos saber si ocurre o no realizando el experimento. Es decir, el que A ocurra o no depende de cual sea el resultado del experimento. Para algunos resultados, la propiedad que define al evento A se presentará y para otros resultados no se presentará. El evento A divide entonces al conjunto de todos los posibles resultados del experimento en dos clases; una clase está formada por todos aquellos resultados de los cuales se deduce la ocurrencia de A y la otra por todos aquellos resultados de los cuales se deduce la no ocurrencia de A. Al conjunto formado por los resultados de la primera clase lo llamaremos el conjunto de resultados favorables a la ocurrencia de A.

DEFINICIÓN 1.20 (**Resultados favorables a un evento**). *Los resultados favorables a un evento A, relativo a un experimento aleatorio, son todos aquellos posibles resultados del experimento de los cuales se deduce la ocurrencia de A.*

De esta manera, cada evento A relativo a un experimento aleatorio puede representarse como un subconjunto del espacio muestral Ω de dicho experimento, el formado por los posibles resultados que le son favorables.

EJEMPLO 1.21. *Consideremos el experimento aleatorio consistente en lanzar* 3 *dados sobre una mesa, anotando los números que se obtienen, y llamemos A al evento: 'se obtiene el mismo número en los* 3 *dados', entonces podemos escribir:*

$$A = \{(1,1,1),(2,2,2),(3,3,3),(4,4,4),(5,5,5),(6,6,6)\}$$

▲

Por otra parte, cada subconjunto A del espacio muestral Ω representa un evento, a saber, el evento: 'ocurre alguno de los resultados que conforman A'[2].

Dado un evento específico, el subconjunto del espacio muestral que lo representa queda únicamente determinado; por otra parte, dado un subconjunto del espacio muestral, se le pueden asociar diferentes eventos, los cuales, formalmente, son distintos; por ejemplo, si consideramos el experimento aleatorio consistente en lanzar un dado 3 veces en forma consecutiva, al subconjunto del espacio muestral $A = \{(1,1,1)\}$ se le pueden asociar los eventos 'la suma de los números que se obtienen es igual a 3' y 'se

[2]En la formulación moderna de la Teoría de la Probabilidad se restringe la familia de eventos de tal manera que se pueda garantizar que la función de probabilidad satisfaga determinadas propiedades. Este punto será analizado más adelante.

obtiene 1 en cada uno de los lanzamientos'. Sin embargo, puede observarse que todos los posibles eventos que pueden asociarse con un subconjunto del espacio muestral tienen la característica de que uno de ellos ocurre en una realización del experimento si y sólo si ocurren los otros. Esto motiva la siguiente definición:

DEFINICIÓN 1.22 (**Equivalencia de eventos**). *Dos eventos son equivalentes si la ocurrencia de cualquiera de ellos implica la ocurrencia del otro en cualquier realización del experimento.*

Con base en esta definición, la familia de eventos queda partida en clases formadas por eventos que son equivalentes entre sí. Todos los eventos de una clase son esencialmente el mismo y así lo consideraremos en lo sucesivo. De esta manera, hay una correspondencia uno a uno entre los eventos relativos a un experimento aleatorio y los subconjuntos del correspondiente espacio muestral.

En particular, cada posible resultado $\omega \in \Omega$ representa un evento, consistente en la ocurrencia de ω. A esta clase particular de eventos, que se representan por un elemento de Ω, los llamaremos eventos elementales.

DEFINICIÓN 1.23 (**Eventos elementales**). *Un evento elemental relativo a un experimento aleatorio es un evento consistente en la ocurrencia de un específico posible resultado del experimento.*

De la misma manera, el mismo espacio muestral Ω representa un evento, el cual tiene la particularidad de ocurrir siempre que se realiza el experimento aleatorio correspondiente, razón por la cual es llamado el evento seguro.

DEFINICIÓN 1.24 (**Evento seguro**). *El evento seguro relativo a un experimento aleatorio es un evento que siempre ocurre al realizar el experimento.*

Finalmente, el conjunto vacío $\emptyset$ representa también un evento, el cual no tiene asociados resultados favorables y que podría definirse como: 'no ocurre ninguno de los posibles resultados del experimento'. Evidentemente este evento nunca ocurre, razón por la cual es llamado el evento imposible.

DEFINICIÓN 1.25 (**Evento imposible**). *El evento imposible relativo a un experimento aleatorio es un evento que nunca ocurre al realizar el experimento.*

1.7. Composición de eventos

Consideremos un experimento aleatorio, digamos lanzar 3 veces un dado en forma consecutiva. Con relación a un experimento de este tipo, hemos dicho que podemos definir eventos, los cuales tienen la característica de ocurrir o no ocurrir cuando realizamos el experimento. En el ejemplo en consideración, los siguientes son eventos:

A: Se obtiene una suma igual a 7.

B: Se obtiene 3 números iguales.

C: Se obtiene por lo menos un 5.

D: Se obtiene una suma igual a 3.

E: En el primer lanzamiento se obtiene 6.

F: En el segundo lanzamiento se obtiene 3.

G: No se obtiene ningún 5.

H: Se obtiene una suma menor que 20.

Estos eventos se pueden comparar unos con otros, por ejemplo, al comparar el evento A con el evento B, podemos decir que, al realizar el experimento, es imposible que ocurran ambos; al comparar el evento C con el evento G, podemos decir que, al realizar el experimento, uno ocurre si y sólo si el otro no ocurre; al comparar el evento E con el evento F, podemos decir que la ocurrencia o no ocurrencia de uno de ellos no influye sobre la ocurrencia o no ocurrencia del otro; comparando los eventos B y D, podemos decir que la ocurrencia de D en una realización del experimento implica la ocurrencia de B.

También a partir de los eventos dados, podemos definir nuevos eventos. Por ejemplo, a partir de C y B podemos considerar un evento definido por la propiedad de que ocurre en la realización de un experimento si y sólo si ocurre alguno de los dos eventos considerados o ambos, en este caso, el nuevo evento así definido ocurre sólo cuando se obtienen 3 números iguales, o bien cuando los 3 números no son iguales pero se obtiene por lo menos un 5; también podemos considerar otro evento definido por la propiedad de que ocurre en la realización de un experimento si y sólo si ocurren ambos eventos, en este caso, el nuevo evento así definido ocurre únicamente cuando se obtienen tres cincos.

Las consideraciones anteriores nos llevan a las siguientes definiciones, para las cuales supondremos que se tiene definido un experimento aleatorio $\mathcal{E}$ y eventos A, B, C,...
relativos a ese experimento.

Definición 1.26 (**Unión de eventos**). *Si A y B son dos eventos, definimos un nuevo evento caracterizado por la propiedad de que ocurre en la realización de un* *experimento si y sólo si ocurre alguno de los eventos A o B, o ambos. A este nuevo evento lo llamaremos la unión de A y B y lo denotaremos por $A \cup B$.*

DEFINICIÓN 1.27 (**Intersección de eventos**). *Si A y B son dos eventos, definimos un nuevo evento caracterizado por la propiedad de que ocurre en la realización de un experimento si y sólo si los dos eventos A y B ocurren. A este nuevo evento lo llamaremos la intersección de A y B y lo denotaremos por $A \cap B$.*

DEFINICIÓN 1.28 (**Complemento de un evento**). *Si A es un evento, definimos un nuevo evento caracterizado por la propiedad de que ocurre en la realización de un experimento si y sólo si A no ocurre. A este nuevo evento lo llamaremos el complemento o la negación de A y lo denotaremos por A^c.*

EJEMPLO 1.29. *Considerando el experimento y los eventos definidos al inicio de esta sección, tenemos:*

$E \cup F$: *se obtiene* 6 *en el primer lanzamiento o se obtiene* 3 *en el segundo lanzamiento, o bien se obtienen ambos resultados.*

$A \cap C$: *Se obtiene un* 5 *y* 2 *unos.*

B^c: *No se obtienen* 3 *números iguales.*

▲

La terminología de la Teoría de Conjuntos utilizada en las definiciones anteriores no es fortuita pues hemos visto que hay una correspondencia uno a uno entre los eventos relativos a un experimento aleatorio y los subconjuntos del correspondiente espacio muestral, de tal manera que las operaciones entre conjuntos se traducen inmediatamente en operaciones entre eventos y resulta evidente que las definiciones anteriores corresponden precisamente a las correspondientes operaciones entre conjuntos. De esta manera, todas las propiedades que se tienen para las operaciones entre conjuntos se traducen inmediatamente en propiedades de las operaciones entre eventos. Entre estas propiedades podemos destacar la conmutatividad, la asociatividad, la distributividad y las leyes de Morgan.

Todas las operaciones entre conjuntos se traducen en operaciones entre eventos, por ejemplo, dados dos eventos A y B, se puede definir la diferencia $A - B$ como el evento $A \cap B^c$, es decir como el evento que ocurre si y sólo si ocurre el evento A pero no el evento B.

De la misma manera, las relaciones que se pueden establecer entre conjuntos se traducen en relaciones entre eventos; por ejemplo, se puede decir que el evento A está contenido en el evento B, y denotarse por $A \subset B$, si, vistos como subconjuntos del espacio muestral, A y B tienen esa relación; es decir si la ocurrencia del evento A implica la ocurrencia del evento B.

EJEMPLO 1.30. *Sean A y B dos eventos, entonces el evento 'ocurre A pero no B' se puede representar por $A \cap B^c$; el evento 'ocurre exactamente uno de los dos eventos A y B' se puede representar por $A \cup B - A \cap B$; el evento 'no ocurre ninguno de los eventos A y B' se puede representar por $A^c \cap B^c$ o bien por $(A \cup B)^c$.*

▲

En la Teoría de Conjuntos se tiene el concepto de eventos ajenos, lo cual significa que éstos tienen una intersección vacía. Como veremos más adelante, este concepto es especialmente importante en la Teoría de la Probabilidad. Cuando dos eventos se representan mediante dos conjuntos ajenos se dice que dichos eventos son mutuamente excluyentes; en otras palabras, se tiene la siguiente definición.

DEFINICIÓN 1.31 (**Eventos mutuamente excluyentes**). *Diremos que dos eventos A y B son mutuamente excluyentes si la ocurrencia de ambos en cualquier realización del experimento es imposible.*

Este concepto, si bien simple, en ocasiones se vuelve confuso al compararlo con el de independencia, el cual se define más adelante (ver sección 2.7). Por tal motivo conviene ilustrarlo aquí con un ejemplo:

EJEMPLO 1.32. *Un experimento aleatorio consiste en lanzar un dado dos veces en forma consecutiva. El resultado del experimento consiste en la pareja ordenada de números que muestran las caras superiores de los dados. Los eventos A y B que se definen a continuación son entonces mutuamente excluyentes:*

A: Se obtiene 6 en el primer lanzamiento.

B: Se obtiene 5 en el primer lanzamiento.

Sin embargo, si C es el evento 'se obtiene 6 en el segundo lanzamiento', los eventos A y C no son mutuamente excluyentes:

Esto es así porque los eventos A y B no pueden, en ningún caso, ocurrir ambos en alguna realización del experimento aleatorio; en cambio, es posible que los eventos A y C ocurran ambos en alguna realización del experimento.

DEFINICIÓN 1.33 (**Eventos mutuamente excluyentes**). *Diremos que n eventos $A_1, \ldots, A_n$ son mutuamente excluyentes si los eventos A_i y A_j son mutuamente excluyentes para toda $i, j \in \{1, \ldots, n\}$, con $i \neq j$.*

Un problema de particular importancia consiste en expresar un evento dado en términos de otros eventos más simples. Por ejemplo, como veremos más adelante, en muchas ocasiones se busca expresar un evento dado en términos de una unión de eventos más simples que el evento dado y de tal manera que estos últimos sean mutuamente excluyentes.

Ejemplo 1.34. *En una cierta compañía el esquema para aprobar una propuesta es el siguiente: tres personas —A, B y C— analizan la propuesta y ésta es aprobada únicamente si por lo menos dos de las tres personas dan su visto bueno. Consideremos una determinada propuesta y sean A, B, C y D los eventos 'la persona A da su visto bueno', 'la persona B da su visto bueno', 'la persona C da su visto bueno' y 'la propuesta es aprobada', respectivamente. D puede expresarse como una unión de eventos mutuamente excluyentes, cada uno de los cuales está dado en términos de A, B y C, a saber, $D = (A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C) \cup (A \cap B \cap C)$.*

1.8. Funciones de probabilidad

Como hemos dicho antes, el problema que se plantea en el Cálculo de Probabilidades consiste en encontrar una manera que permita asignar a cada evento un número, el cual mida el "que tan fácilmente" se presentará el evento en futuras realizaciones de un experimento aleatorio. Ese número que se asigna a un evento es llamado, como ya dijimos también, la probabilidad del evento. Dicho de otra manera, lo que se busca es una función con valores reales definida sobre la familia de eventos; a tal función la llamaremos la **función de probabilidad**.

Al asignar probabilidades a cada evento, relativo a un experimento aleatorio dado, lo que estamos haciendo es definir un modelo que nos permitirá estudiar el experimento aleatorio. En un problema real cada probabilidad asignada tendrá una interpretación práctica y con base en esa interpretación se podrá decidir sobre la fidelidad con que el modelo representa al fenómeno aleatorio en consideración. La asignación de probabilidades es así, en un cierto sentido, arbitraria, pues se puede proponer una cierta asignación, desarrollar el modelo con base en ella y más adelante, de acuerdo a la interpretación práctica que demos a la probabilidad, decidir sobre la validez del modelo como representación del experimento aleatorio. Sin embargo, como veremos en los ejemplos desarrollados en este libro, para nosotros la probabilidad de un evento refleja siempre determinadas características del experimento aleatorio en consideración, es decir, tiene un sentido objetivo. De esta manera, al desarrollar un modelo para un experimento aleatorio buscaremos que la asignación de probabilidades no sea arbitraria, debiendo reflejar de alguna manera las características del experimento aleatorio.

La interpretación práctica de la probabilidad de un evento que se utiliza más frecuentemente es la basada en el principio de regularidad de las frecuencias. Según esta interpretación, la probabilidad de un evento A, relativo a un experimento aleatorio repetible, es un número que debe ser aproximadamente igual a la frecuencia relativa con que el evento A ocurre en una serie grande de realizaciones del experimento aleatorio. A lo largo de este texto, utilizaremos esta interpretación en diferentes situaciones, en particular, en lo que sigue, la usaremos para deducir algunas reglas que impondremos sobre nuestros modelos.

Si interpretamos a la probabilidad de un evento de acuerdo con el principio de regularidad de las frecuencias, se imponen algunas restricciones sobre el modelo que queremos construir. En particular, siendo la frecuencia relativa con que ocurre un evento una cantidad no negativa, la probabilidad de un evento cualquiera debe tener entonces la misma propiedad. En otras palabras, podemos escribir la siguiente regla:

Si A es un evento y denotamos por $P(A)$ a su probabilidad de ocurrencia, entonces $P(A) \geq 0$.

Por otro lado, el evento seguro Ω tiene, en cualquier caso, una frecuencia relativa igual a 1 pues siempre ocurre. Es decir, podemos escribir $P(\Omega) = 1$.

Supongamos ahora que un cierto evento C es la unión de dos eventos mutuamente excluyentes A y B y que hemos logrado de alguna manera asignar las probabilidades $P(A)$ y $P(B)$ a los eventos A y B respectivamente. De acuerdo a la interpretación que estamos dando a la probabilidad de un evento, $P(A)$ y $P(B)$ representan las frecuencias relativas con que ocurren los eventos A y B, respectivamente, en una serie grande de realizaciones del experimento aleatorio. Resulta entonces claro que la suma $P(A) + P(B)$ representa la frecuencia relativa con que ocurre el evento C y se tiene entonces $P(C) = P(A) + P(B)$.

Hemos derivado las 3 propiedades anteriores del supuesto de que podemos interpretar a la probabilidad de un evento como un número que debe ser aproximadamente igual a la frecuencia relativa con que el evento ocurre en una serie grande de realizaciones del experimento aleatorio. Sin embargo, las propiedades mismas no involucran el uso de la frecuencia relativa con que ocurre cada uno de los eventos en consideración. Más aún, si pensamos a la función de probabilidad que buscamos como una medida definida sobre los eventos relativos al experimento aleatorio en consideración, las propiedades primera y tercera pueden considerarse como las mínimas que se pueden pedir, mientras que la segunda puede verse simplemente como una normalización de dicha medida. En otras palabras, esas 3 propiedades parecen ser lo suficientemente razonables como para ser compatibles con cualquier interpretación práctica que se dé a la probabilidad. Con base en esto, tomaremos esas propiedades como punto de partida en la construcción del modelo que estamos buscando. En otras palabras, se tiene la siguiente definición:

DEFINICIÓN 1.35 (**Función de probabilidad**). *Dado un experimento aleatorio cualquiera, denotando por $\mathcal{A}$ a la familia de eventos relativos a ese experimento, diremos* *que una función $P : \mathcal{A} \mapsto \mathbb{R}$ es una función de probabilidad si se satisfacen las siguientes propiedades:*

(i) $P(A) \geq 0$ *para todo evento* A.
(ii) $P(\Omega) = 1$.
(iii) *Si* A *y* B *son dos eventos mutuamente excluyentes, entonces:*
$$P(A \cup B) = P(A) + P(B)$$

EJERCICIOS

Ejercicio 1.1 (**Problema de los 3 jugadores**). *Tres jugadores —P,Q y R— juegan partidas por parejas, comenzando P contra Q. Quien gane una partida juega con el otro jugador, hasta que uno de los jugadores gane dos partidas consecutivas, ganando entonces el juego. Describa el espacio muestral de este juego.*

Ejercicio 1.2. *Un experimento aleatorio consiste en elegir al azar* 3 *números —a, b y c— en el intervalo* $[0, 1]$*, para formar la ecuación* $ax^2 + bx + c = 0$*. Sea* A *el evento 'las dos raíces de la ecuación son reales y distintas'. Represente como un conjunto* Ω *al espacio muestral de este experimento y al evento* A *como un subconjunto de* Ω.

Ejercicio 1.3. *Dos personas, P y Q, juegan un juego de azar, el cual consiste en ir lanzando un par de dados por turnos, comenzando por P, de tal manera que, si P obtiene una suma igual a* 7*, se acaba el juego, ganando P, mientras que, si Q obtiene una suma igual a* 6*, se acaba el juego, ganando Q. Sea* A *el evento 'el jugador P gana el juego'. Represente como un conjunto* Ω *al espacio muestral de este juego y al evento* A *como un subconjunto de* Ω.

Ejercicio 1.4. *Sean* A *y* B *dos eventos y sean* E *y* F *los eventos 'ocurre exactamente uno de los dos eventos* A *y* B*' y 'ocurre a lo más uno de los dos eventos* A *y* B*', respectivamente. Exprese los eventos* E *y* F *en términos de* A *y* B.

Ejercicio 1.5. *Un trabajador produce* n *partes de un artículo. Sea* A_i *el evento 'la* i*-ésima parte está defectuosa'. Exprese en términos de los* A_i *cada uno de los siguientes eventos: a) ninguna de las* n *partes está defectuosa, b) al menos una de las* n *partes está defectuosa, c) exactamente una de las* n *partes está defectuosa.*

CAPÍTULO 2

LAS REGLAS BÁSICAS

Mientras más difícil parezca determinar por la razón lo que es incierto y está sometido al azar, la ciencia que logre ese resultado parecerá más admirable.

Christiaan Huygens

2.1. Algunas propiedades elementales

De las 3 propiedades que estamos pidiendo a cualquier función de probabilidad se derivan otras, entre las que destacan las siguientes:

PROPOSICIÓN 2.1. *Si A es cualquier evento, entonces $P(A) \leq 1$ y $P(A^c) = 1 - P(A)$.*

Demostración

Los eventos A y A^c son mutuamente excluyentes y su unión es el evento seguro Ω, así que $1 = P(\Omega) = P(A) + P(A^c)$, de lo cual se sigue el resultado.

■

PROPOSICIÓN 2.2. $P(\emptyset) = 0$.

Demostración

Siendo Ω y $\emptyset$ complementarios, se tiene, por la proposición anterior, $P(\emptyset) = 1 - P(\Omega) = 1 - 1 = 0$.

■

PROPOSICIÓN 2.3. *Si A y B son eventos tales que $A \subset B$ entonces $P(B - A) = P(B) - P(A)$.*

Demostración

Se tiene $B = A \cup (B - A)$ y los eventos A y $B - A$ son mutuamente excluyentes, de manera que se tiene, $P(B) = P(A) + P(B - A)$, de lo cual se sigue el resultado.

∎

PROPOSICIÓN 2.4. *Si A y B son eventos tales que $A \subset B$, entonces $P(A) \leq P(B)$.*

Demostración

Por la proposición anterior, se tiene $P(B) - P(A) = P(B - A)$, de manera que, siendo $P(B - A)$ un número real no negativo, se tiene el resultado.

∎

2.2. Propiedad de la aditividad finita

Una propiedad especialmente importante de la función de probabilidad está dada en la siguiente proposición:

PROPOSICIÓN 2.5 (**Propiedad de la aditividad finita**). *Sean $A_1, \ldots, A_n$ n eventos mutuamente excluyentes, entonces, $P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$.*

Demostración

La demostración es por inducción sobre el número de eventos n. Para $n = 2$ el resultado se obtiene directamente de la definición 1.35.

Supongamos ahora que la propiedad es válida para el caso de cualesquiera $n-1$ eventos mutuamente excluyentes y sean $A_1, \ldots, A_n$ n eventos mutuamente excluyentes. Los eventos $\bigcup_{k=1}^{n-1} A_k$ y A_n son entonces mutuamente excluyentes, así que se tiene:

$P(\bigcup_{k=1}^{n} A_k) = P(\bigcup_{k=1}^{n-1} A_k) + P(A_n) = \sum_{k=1}^{n-1} (PA_k) + P(A_n) = \sum_{k=1}^{n} (PA_k)$

∎

La propiedad de la aditividad finita puede considerarse como la propiedad básica de la función de probabilidad pues ésta permite establecer un método para calcular probabilidades de eventos. Mediante este método, dado un evento A, cuya probabilidad se busca, se trata de encontrar una descomposición de A en una unión de eventos mutuamente excluyentes cuyas probabilidades sean más simples de calcular; entonces la probabilidad de A se obtiene como la suma de esas probabilidades. Numerosos ejemplos de esta situación se presentarán a lo largo del texto.

De manera general, **el método para asignar probabilidades a los eventos relativos a cualquier experimento aleatorio va de lo simple a lo complejo:**

primero se encuentra la probabilidad de una clase particular de eventos y, a partir de ahí, utilizando las propiedades de la función de probabilidad, se extiende ésta a una clase más amplia de eventos y después a familias cada vez más extensas. Éste será el método que utilizaremos a los largo de este texto.

PROPOSICIÓN 2.6 (**Subaditividad finita o desigualdad de Boole**). *Sean $A_1, \ldots A_n$ n eventos, entonces, $P\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} P(A_k)$.*

Demostración

Sea $B_1 = A_1$ y, para $k \in \{2, \ldots, n\}$, $B_k = A_k - \bigcup_{j=1}^{k-1} A_j$, entonces $\bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{n} B_k$, $B_k \subset A_k$ para cualquier $k \in \{1, \ldots, n\}$ y los eventos $B_1, \ldots, B_n$ son mutuamente excluyentes, así que:

$P\left(\bigcup_{k=1}^{n} A_k\right) = P\left(\bigcup_{k=1}^{n} B_k\right) = \sum_{k=1}^{n} P(B_k) \leq \sum_{k=1}^{n} P(A_k)$

■

2.3. Regla de la suma

Al igual que la propiedad de la aditividad finita, la siguiente propiedad generaliza la propiedad *iii* de la función de probabilidad y es también particularmente importante:

PROPOSICIÓN 2.7 (**Regla de la suma para 2 eventos**). *Si A y B son dos eventos cualesquiera, entonces:*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Demostración

Se tiene $A \cup B = A \cup (B - A \cap B)$ y los eventos A y $B - A \cap B$ son mutuamente excluyentes, así que se tiene, $P(A \cup B) = P(A) + P(B - A \cap B) = P(A) + P(B) - P(A \cap B)$.

■

La siguiente proposición, cuya demostración se deja como ejercicio, generaliza la proposición 2.7 para el caso de n eventos cualesquiera.

PROPOSICIÓN 2.8 (**Regla de la suma para n eventos**). *Sean $A_1, \ldots, A_n$ n eventos cualesquiera, entonces:*

$P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k) - \sum_{\{i,j \in \{1,\ldots,n\},: i \neq j\}} P(A_i \cap A_j)$

$+ \sum_{\{i,j,k \in \{1,\ldots,n\},: i \neq j, j \neq k, i \neq k\}} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P(A_1 \cap A_2 \ldots \cap A_n)$

Ejemplo 2.9. *Dos eventos A y B son tales que $P(A) = 0.3$, $P(B) = 0.4$ y $P(A \cap B) = 0.1$. Encuentre la probabilidad de que a) ocurra exactamente uno de los dos eventos A y B y b) no ocurra ninguno de los dos eventos.*

Solución

a. $P(A - A \cap B) + P(B - A \cap B) = 0.3 - 0.1 + 0.4 - 0.1 = 0.5$

b. $P(A^c \cap B^c) = P\left((A \cup B)^c\right) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A \cap B)$

$= 1 - 0.3 - 0.4 + 0.1 = 0.4$

▲

Para implementar el método para asignar probabilidades, descrito en la sección 2.2, se requiere obviamente de una manera que permita iniciar el proceso, lo cual es objeto de parte del siguiente tema. Además, con el objeto de tener todas las herramientas que permiten extender la función de probabilidad, se tratarán también los temas de la probabilidad condicional y otros relacionados.

2.4. Elecciones al azar y resultados equiprobables

En una clase bastante amplia de problemas, la asignación de probabilidades a los eventos de un experimento aleatorio puede iniciarse gracias al concepto de **"elección al azar"**.

La elección al azar se refiere a experimentos aleatorios en los cuales se dispone de una colección de objetos, los cuales pueden ser bolas, cajas, tarjetas, personas, etc., y se define el experimento aleatorio precisamente como la elección al azar de uno o varios objetos de la colección. El término "elegir al azar" que se utiliza en esos casos se entiende en el sentido de que, al elegir los objetos, no se tiene preferencia por la elección de uno de ellos sobre la elección de otro. Ahora bien, si recordamos que la probabilidad de un evento es un número que representa la facilidad con que el evento se presenta al realizar el experimento aleatorio, entonces, al considerar el experimento aleatorio consistente en "elegir al azar" uno o varios objetos de la colección, los posibles resultados del experimento deberán tener asignada la misma probabilidad, en cuyo caso diremos que los posibles resultados son equiprobables. De aquí que, en realidad, el término "elección al azar" es sinónimo de equiprobabilidad de los posibles resultados de la elección.

Un experimento aleatorio puede estar compuesto por dos o más experimentos definidos como elecciones al azar y en algunos casos la equiprobabilidad se da únicamente parcialmente, pero, de cualquier manera, esto puede ser suficiente para determinar la probabilidad de cada uno de los posibles resultados del experimento compuesto. Esto resuelve nuestro problema pues una vez asignada una probabilidad a cada evento elemental, la probabilidad de cualquier evento compuesto por un número finito de eventos elementales se puede calcular, gracias a la propiedad de la aditividad finita, simplemente sumando las probabilidades de los eventos elementales que lo componen.

En particular, aunque no siempre es el caso, los posibles resultados de un experimento aleatorio pueden resultar equiprobables. En ese caso resulta aún más simple asignar una probabilidad a cualquier evento. En primer lugar, con base en la propiedad *ii* de la función de probabilidad y en la propiedad de la aditividad finita, la suma de las probabilidades asignadas a los eventos elementales debe ser igual a 1, de manera que si son N los posibles resultados del experimento y todos ellos tienen asignada la misma probabilidad, cada una de ellas debe ser entonces igual a $\frac{1}{N}$. Sea entonces $\Omega = \{\omega_1, \ldots, \omega_N\}$ el espacio muestral del experimento aleatorio y $A = \{\omega_{j_1}, \ldots, \omega_{j_n}\}$ un evento cualquiera; por la propiedad de la aditividad finita, la probabilidad del evento A está dada por $P(A) = \frac{n}{N}$.

En otras palabras, en este caso, para calcular la probabilidad de un evento A, es necesario únicamente encontrar el número de eventos elementales que lo componen y dividir este número por el número total de eventos elementales.

Este método de encontrar probabilidades, basándose en la equiprobabilidad de los posibles resultados, es conocido como la **definición clásica de probabilidad**. Según esta definición, la probabilidad de un evento A se calcula de la siguiente manera:

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

relación que es exactamente la misma que la que hemos dado anteriormente.

Debe tenerse presente que esta manera de calcular probabilidades puede aplicarse únicamente cuando se ha determinado que los posibles resultados del experimento aleatorio son equiprobables.

En algunos casos se tienen experimentos aleatorios en donde intervienen "simetrías", las cuales hacen que tales experimentos se puedan considerar equivalentes a una elección al azar. Un ejemplo típico de esa situación se presenta al considerar el experimento consistente en lanzar un dado sobre una mesa y observar el número que muestra finalmente su cara superior. Si se trata de un dado simétrico y balanceado, por consideraciones de simetría, este experimento puede considerarse equivalente a seleccionar al azar un número del conjunto $\{1, 2, 3, 4, 5, 6\}$, de tal manera que la probabilidad de cada uno de sus posibles resultados es igual a $\frac{1}{6}$.

EJEMPLO 2.10. *Un experimento aleatorio consiste en seleccionar al azar un paquete de dos bolas de una urna, la cual contiene* 4 *bolas rojas y* 6 *bolas blancas. ¿Cuál es la probabilidad de que las dos bolas seleccionadas sean de color rojo?*

Solución

Dado que la elección es al azar y que el número de paquetes de dos bolas que se pueden obtener como resultado del experimento es igual a 45*, se puede concluir que la obtención de cada uno de esos paquetes es igual a* $\frac{1}{45}$*. Por otra parte, de los* 45 *posibles paquetes de dos bolas, hay* 6 *en los cuales las dos bolas son rojas. Por lo tanto, la probabilidad que se busca es igual a* $\frac{6}{45}$*.*

▲

Como ya dijimos, el concepto de elección al azar puede aplicarse de manera parcial y esto ser suficiente para determinar la función de probabilidad. Como ilustración consideremos el siguiente ejemplo:

EJEMPLO 2.11. *Una primera urna contiene* 2 *bolas rojas y* 4 *blancas; una segunda urna contiene* 8 *bolas rojas y* 7 *blancas; y una tercera urna contiene* 6 *bolas rojas y* 4 *blancas. Un experimento aleatorio consiste de dos partes, en la primera parte se selecciona una urna al azar y en la segunda parte se selecciona al azar una bola de la urna elegida en la primera parte. ¿Cuál es la probabilidad de que se seleccione una bola roja?*[1]

Solución

El espacio muestral de este experimento aleatorio tiene 31 *elementos, uno por cada una de las posibles bolas que se seleccionan. Representemos por* R_1, R_2 *las bolas rojas de la primera urna; por* $R_3, R_4, R_5, R_6, R_7, R_8, R_9, R_{10}$ *las bolas rojas de la segunda urna, etc. De la misma manera, representemos las bolas blancas de la primera urna por* B_1, B_2, B_3, B_4*, etc., y consideremos, para* $i \in \{1, 2, 3\}$*, los siguientes eventos:*

A_i*: Se elige la* i*-ésima urna.*

B*: Se selecciona una bola roja.*

Como la elección de la urna se realiza al azar, se tiene, $P(A_i) = \frac{1}{3}$*. Por otra parte, dado que, una vez seleccionada la urna, la elección de la bola se realiza también al azar, la probabilidad de seleccionar cualquiera de las bolas de la primera urna es la misma para todas ellas; lo mismo puede decirse de la probabilidad de seleccionar*

[1]En realidad en este problema está implícito el concepto de probabilidad condicional, el cual se introduce más adelante. Se trata aquí únicamente para ilustrar el hecho de que el concepto de elección al azar permite encontrar la función de probabilidad, incluso en casos en donde los posibles resultados del experimento no son equiprobables.

cualquiera de las bolas de la segunda o de la tercera urna (sin embargo, la probabilidad de seleccionar cualquiera de las 31 *bolas no es la misma para todas ellas). Llamemos* p_i *a la probabilidad común de seleccionar cualquiera de las bolas de la i-ésima urna. Los eventos* A_i *tienen la siguiente representación:*

$A_1 = \{R_1, R_2, B_1, B_2, B_3, B_4\}$

$A_2 = \{R_3, R_4, R_5, R_6, R_7, R_8, R_9, R_{10}, B_5, B_6, B_7, B_8, B_9, B_{10}, B_{11}\}$

$A_3 = \{R_{11}, R_{12}, R_{13}, R_{14}, R_{15}, R_{16}, B_{12}, B_{13}, B_{14}, B_{15}\}$

Así que, aplicando la propiedad de la aditividad finita, se tiene:

$\frac{1}{3} = P(A_1) = 6p_1$

$\frac{1}{3} = P(A_2) = 15p_2$

$\frac{1}{3} = P(A_3) = 10p_3$

Por lo tanto, $p_1 = \frac{1}{18}$, $p_2 = \frac{1}{45}$ *y* $p_3 = \frac{1}{30}$.

Finalmente, aplicando nuevamente la propiedad de la aditividad finita, se tiene:

$P(B) = \frac{2}{18} + \frac{8}{45} + \frac{6}{30} = \frac{22}{45}$

Una aplicación general de esta idea se realizará posteriormente en el estudio del muestreo aleatorio.

2.5. Probabilidad condicional

En ocasiones un experimento aleatorio está compuesto por varios experimentos, también aleatorios, los cuales se realizan uno después del otro. Un ejemplo típico de esta situación se obtiene al considerar n elecciones sucesivas, todas ellas al azar, de elementos de una determinada población. El resultado de cada parte puede ser independiente o depender de las partes anteriores. En general en estos casos lo que se tiene a la mano para iniciar el proceso de calcular probabilidades son probabilidades de eventos relativos a cada parte del experimento, condicionadas a lo que haya ocurrido en las primeras partes. Este tipo de probabilidades es conocido con el nombre de **probabilidades condicionales**. Como ilustración, consideremos el ejemplo siguiente:

Ejemplo 2.12. *Una urna contiene* 3 *bolas rojas y* 6 *bolas negras. Un experimento aleatorio consiste en dos partes, en la primera, se selecciona al azar una bola de la* *urna y se deja fuera de ésta, en la segunda, se selecciona al azar una de las bolas restantes en la urna.*

Denotemos por $B_1, B_2, B_3, B_4, B_5, B_6, B_7, B_8, B_9$ *las* 9 *bolas de la urna. Cada posible resultado del experimento aleatorio puede entonces representarse mediante una pareja* (B_i, B_j), *con* $i \neq j$, *en donde* B_i *representa el resultado de la primera elección y* B_j *el de la segunda. El espacio muestral consta así de* 72 *elementos.*

La descripción del experimento no da directamente la probabilidad de cada uno de estos posibles resultados, en cambio, se obtiene inmediatamente la siguiente información:

Como la elección de la primera bola se realiza al azar, cualquiera de las bolas en la urna tiene la misma probabilidad de ser elegida, de manera que si, para $i \in \{1, \ldots, 9\}$, *definimos los siguientes eventos:*

B_i^1: *En la primera elección se obtiene la bola* B_i,

se tiene:

$B_i^1 = \{(B_i, B_1), (B_i, B_2), \ldots (B_i, B_{i-1}), (B_i, B_{i+1}), \ldots (B_i, B_9)\}$

y $P(B_i^1) = \frac{1}{9}$ *para toda* i.

Por otra parte, una vez seleccionada la primera bola, la elección de la segunda se realiza también al azar, de manera que todas las bolas restantes tienen la misma probabilidad de ser elegidas. Esto significa que, para cada $i \in \{1, \ldots, 9\}$, *las probabilidades de cada uno de los* 8 *posibles resultados que componen el evento* B_i^1, *son iguales.*

Con esta información se puede obtener la probabilidad de cada uno de los posibles resultados del experimento pues siendo 8 *los elementos que componen cada* B_i^1, *teniendo todos ellos la misma probabilidad y la suma de ellas siendo igual a* $\frac{1}{9}$, *se concluye que la probabilidad de cada uno de ellos es* $\frac{1}{72}$.

▲

Las probabilidades relacionadas con la segunda parte del experimento del último ejemplo pueden calificarse como condicionales pues se obtienen condicionadas a lo que ocurrió en la primera parte. Podríamos escribir éstas de la siguiente manera:

$$P(B_j^2 | B_i^1) = \tfrac{1}{8} \text{ para } j \in \{1, \ldots, 9\},\ j \neq i$$

en donde B_j^2 representa al evento 'en la segunda elección se obtiene la bola B_j'.

Las probabilidades así definidas se leen: la probabilidad condicional del evento B_j^2 dada la ocurrencia del evento B_i^1.

Lo que hemos visto en el ejemplo anterior es que a partir de las probabilidades $P(B_i^1)$ y $P(B_j^2|B_i^1)$ es posible obtener las probabilidades $P(B_i^1 \cap B_j^2) = P((B_i, B_j))$.

A continuación se muestra como estas ideas pueden extenderse a una situación general y obtener así una herramienta más que nos permitirá extender la función de probabilidad.

Una probabilidad condicional tiene un sentido muy claro en los casos en que se trata de la probabilidad de un evento que depende de una segunda parte de un experimento, condicionada a lo que ocurra en la primera parte de éste. En el caso general de cualquier experimento aleatorio, es posible definir el concepto de probabilidad condicional de un evento B dada la ocurrencia de un evento A, pero cambiando el enfoque.

Decir que un evento A ocurre equivale a decir que ocurre alguno de los posibles resultados que lo componen; en otras palabras los posibles resultados del experimento aleatorio no son ya todos los elementos del espacio muestral sino únicamente aquellos que conforman el evento A. También se puede dar la misma idea diciendo que, cuando el evento A ocurre, el espacio muestral se reduce y queda conformado exclusivamente por los posibles resultados que componen al evento A. La pregunta es: ¿qué relación existe entre la probabilidad de un evento B, calculada con relación a todo el espacio muestral, con la probabilidad del evento B, calculada con relación al espacio muestral restringido? Obviamente estas probabilidades no tienen que ser iguales pues, por ejemplo, dado que A ocurre, la ocurrencia de A se vuelve segura, es decir, se tiene $P(A|A) = 1$; pero A puede tener cualquier otra probabilidad, tomándola con relación a todo el espacio muestral.

Dado que el evento A ocurre, un evento B ocurre si y sólo si ocurre $A \cap B$, es decir, se debe de tener $P(B \mid A) = P(A \cap B \mid A)$, de manera que el problema se reduce a determinar cuál es la probabilidad de un subconjunto del espacio muestral restringido, con relación a éste.

De acuerdo con el significado que queremos darle a la probabilidad de un evento, como una medida del que tan fácilmente ocurre éste al realizar el experimento aleatorio, podemos asumir que las probabilidades en el espacio muestral restringido deben asignarse de tal manera que entre ellas guarden la misma proporción que la que guardan en el espacio muestral original. En otras palabras, si D_1 y D_2 son eventos cualesquiera contenidos en A entonces se debe tener:

$\frac{P(D_2)}{P(D_1)} = \frac{P(D_2|A)}{P(D_1|A)}$

Es decir:

$\frac{P(D_2)}{P(D_2|A)} = \frac{P(D_1)}{P(D_1|A)} = c$

en donde c depende exclusivamente del evento A.

Por otra parte, sabemos que $P(A|A) = 1$, de lo cual se obtiene $c = P(A)$ y, entonces, $P(D \mid A) = \frac{P(D)}{P(A)}$, para cualquier evento $D \subset A$.

Además, $P(B \mid A) = P(A \cap B \mid A)$, para cualquier evento B.

Así, podemos concluir que la siguiente definición es la adecuada:

Definición 2.13 (**Probabilidad condicional**). *Sean A y B dos eventos y supongamos $P(A) > 0$, se define la probabilidad condicional de B, dada la ocurrencia de A, $P(B|A)$, mediante la fórmula:*

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Siendo la probabilidad condicional una probabilidad calculada en un espacio muestral reducido, es de esperarse que ésta tenga las mismas propiedades que cualquier función de probabilidad. En efecto, esto es así, como se demuestra en seguida, de manera que la probabilidad condicional satisface las mismas propiedades que se obtienen como corolarios para cualquier función de probabilidad, por ejemplo, la propiedad de la aditividad finita, la regla de la suma, etc.

Proposición 2.14. *Dado un evento A de probabilidad positiva, la función que asigna a cada evento B el número real $P(B \mid A)$, de acuerdo con la definición 2.13, es una función de probabilidad, de acuerdo con la definición 1.35.*

Demostración

La función así definida es claramente no negativa, además:

$P(\Omega \mid A) = \frac{P(A\cap\Omega)}{P(A)} = \frac{P(A)}{P(A)} = 1$

Si B y C son eventos mutuamente excluyentes, entonces:

$P(B \cup C \mid A) = \frac{P[A\cap(B\cup C)]}{P(A)} = \frac{P[(A\cap B)\cup(A\cap C)]}{P(A)} = \frac{P(A\cap B)+P(A\cap C)}{P(A)}$

$= \frac{P(A\cap B)}{P(A)} + \frac{P(A\cap C)}{P(A)} = P(B \mid A) + P(C \mid A)$

■

EJEMPLO 2.15. *Una urna contiene tarjetas numeradas del* 1 *al* n*, de tal manera que, al seleccionar una de ellas al azar, la probabilidad de obtener un número dado es proporcional a su magnitud. Si se selecciona una tarjeta al azar, ¿cuál es la probabilidad de que esté marcada con el número* 1 *en cada uno de los dos casos siguientes: a) sin disponer de información adicional y b) sabiendo que la bola seleccionada está marcada con algún número entre* 1 *y* m $(1 \leq m \leq n)$*?*

Solución

a. Consideremos, para cada $k \in \{1, \ldots, n\}$*, al evento* A_k*: la tarjeta seleccionada está marcada con el número* k*. Se sabe que* $P(A_k) = ck$*, en donde* c *es una constante; además,* $\sum_{k=1}^{n} P(A_k) = 1$*, es decir,* $c \sum_{k=1}^{n} k = 1$*, así que* $c = \frac{2}{n(n+1)}$ *y, entonces,* $P(A_k) = \frac{2}{n(n+1)}k$*, de manera que,* $P(A_1) = \frac{2}{n(n+1)}$*.*

b. Consideremos, para cada $m \in \{1, \ldots, n\}$*, al evento* B_m*: la tarjeta está marcada con algún número entre* 1 *y* m *(inclusive). Se tiene:*

$P(B_m) = \sum_{k=1}^{m} P(A_k) = \frac{2}{n(n+1)} \sum_{k=1}^{m} k = \frac{m(m+1)}{n(n+1)}$

Así que:

$P\left(A_1 \mid B_m\right) = \frac{P(A_1 \cap B_m)}{P(B_m)} = \frac{P(A_1)}{P(B_m)} = \frac{2}{m(m+1)}$

Obsérvese que el resultado corresponde a la idea de que una probabilidad condicional consiste en una probabilidad calculada en un espacio muestral reducido, a saber, el que corresponde a la condición. El saber que la tarjeta que se obtiene está marcada con algún número entre 1 *y* m *es equivalente a decir que lo que se tiene es una urna con tarjetas numeradas del* 1 *al* m *de tal manera que, al seleccionar una de ellas al azar, la probabilidad de obtener un número dado es proporcional a su magnitud.*

EJEMPLO 2.16. *Sea* C *un evento de probabilidad positiva y* A*,* B *dos eventos tales que, dado que* C *ocurre, la probabilidad de que* A *ocurra es* $\frac{2}{9}$*, la probabilidad de que* B *ocurra es* $\frac{3}{9}$ *y la probabilidad de que no ocurra ni* A *ni* B *es* $\frac{5}{9}$*. Encuentre la probabilidad de que a) ocurra al menos uno de los dos eventos* A *y* B*, b) ocurran los dos eventos* A *y* B *y c) ocurra exactamente uno de los dos eventos* A *y* B*, sabiendo, en los* 3 *casos, que* C *ocurre.*

Solución

a. $P(A \cup B \mid C) = 1 - P\left((A \cup B)^c \mid C\right) = 1 - P\left(A^c \cap B^c \mid C\right) = 1 - \frac{5}{9} = \frac{4}{9}$

b. $P(A \cap B \mid C) = P(A \mid C) + P(B \mid C) - P(A \cup B \mid C) = \frac{2}{9} + \frac{3}{9} - \frac{4}{9} = \frac{1}{9}$

c. $P(A - A \cap B \mid C) + P(B - A \cap B \mid C) = \frac{2}{9} - \frac{1}{9} + \frac{3}{9} - \frac{1}{9} = \frac{3}{9}$

Ejemplo 2.17. *Una cierta fruta crece en dos regiones X y Y, las cuales en ocasiones están infectadas con una plaga que puede dañar la fruta. Es sabido que la probabilidad de que la región X esté infectada es igual a* $\frac{2}{5}$*, la probabilidad de que la región Y lo esté es igual a* $\frac{3}{4}$ *y la probabilidad de que alguna de las dos regiones esté infectada es igual a* $\frac{4}{5}$*. Dado que un envío de fruta que viene de la región X está infectado con la plaga, ¿cuál es la probabilidad de que la región Y también se encuentre infectada?*

Solución

Si llamamos A y B a los eventos 'la región X está infectada con la plaga' y 'la región Y está infectada con la plaga', respectivamente, se sabe $P(A) = \frac{2}{5}$, $P(B) = \frac{3}{4}$ *y* $P(A \cup B) = \frac{4}{5}$*, así que:*

$P(A \cap B) = P(A) + P(B) - P(A \cup B) = \frac{2}{5} + \frac{3}{4} - \frac{4}{5} = \frac{7}{20}$

Por lo tanto:

$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{7}{20}}{\frac{2}{5}} = \frac{7}{8}$

2.6. Regla del producto

La definición de probabilidad condicional puede utilizarse en dos sentidos diferentes; por un lado, se puede utilizar de manera directa para calcular la probabilidad de un evento B, dado que ocurre algún otro evento A, es decir, con respecto a un espacio muestral reducido; por otro lado, la definición puede expresarse en la forma siguiente:

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

relación que es conocida como la **regla del producto** (para dos eventos).

Esta forma, si bien expresa exactamente la misma relación que la forma original, se aplica en un sentido muy distinto pues permite calcular la probabilidad de una intersección de eventos en aquellos casos en que la probabilidad condicional de la fórmula puede obtenerse de manera directa. Esta idea ya la aplicamos en el ejemplo 2.12 en el cual las condiciones del problema determinaban algunas probabilidades no condicionales y otras condicionales; a partir de ambas, se pudo establecer la probabilidad de ocurrencia de cada uno de los posibles resultados del experimento aleatorio, los cuales consisten, en ese caso, de intersecciones de eventos.

La idea general que puede establecerse aquí es la siguiente:

Dado un determinado experimento aleatorio, se trata de establecer, con base en las características del experimento, las probabilidades (condicionales o no) de una clase particular de eventos. Una vez establecidas éstas, el problema radica en extender la función de probabilidad a una clase de eventos tan amplia como sea posible, utilizando las propiedades que tiene la función de probabilidad. Las probabilidades condicionales que puedan asignarse en la primera parte de este proceso no se calculan con base en la definición, sino que más bien se establecen como características que se imponen al modelo matemático que estamos construyendo para el experimento aleatorio en consideración. En la segunda parte del proceso, las propiedades básicas de la función de probabilidad, las cuales permiten su extensión, son la regla de la suma y la del producto.

La regla del producto, establecida para dos eventos, puede generalizarse al caso de n eventos, obteniéndose así la forma general de la regla del producto, la cual, junto a la forma general de la regla de la suma, constituye, como ya lo mencionamos, una de las herramientas básicas en el Cálculo de Probabilidades.

PROPOSICIÓN 2.18 (**Regla del producto**). *Sean $A_1, \ldots, A_n$ n eventos tales que $P(A_1 \cap \ldots \cap A_{n-1}) > 0$, entonces:*

$$P(\cap_{k=1}^{n} A_k) = P(A_n | A_1 \cap \ldots \cap A_{n-1}) \cdots P(A_2 | A_1) P(A_1)$$

La demostración de esta proposición se deja como ejercicio.

EJEMPLO 2.19. *Una urna contiene 15 bolas negras y 10 bolas blancas. Un experimento aleatorio consiste en seleccionar al azar, una por una, 10 bolas de la urna, de tal manera que cada bola seleccionada se deja fuera de la urna. Calcule la probabilidad de que la primera bola blanca se obtenga en la quinta elección.*

Solución

Sea N_k el evento 'se obtiene bola negra en la k-ésima elección' y B_k el evento 'se obtiene bola blanca en la k-ésima elección', entonces:

$P(N_1 \cap N_2 \cap N_3 \cap N_4 \cap B_5)$

$= P(B_5 \mid N_1 \cap N_2 \cap N_3 \cap N_4) P(N_4 \mid N_1 \cap N_2 \cap N_3) \cdots P(N_1)$

$= \frac{10}{21} \frac{12}{22} \frac{13}{23} \frac{14}{24} \frac{15}{25} = \frac{13}{253} = 0.051383$

EJEMPLO 2.20. *Se van colocando al azar bolas, una a una, en cualquiera de N cajas, hasta que alguna caja llegue a tener 2 bolas. ¿Cuál es la probabilidad de que el proceso termine en el paso n, en donde $2 \leq n \leq N+1$?*

Solución

Consideremos los siguientes eventos:

A*: el proceso termina en el paso* n*.*

B_k*: la* k*-ésima bola que se coloca queda en una caja desocupada.*

B*: la* n*-ésima bola que se coloca queda en una caja ocupada.*

Se tiene:

$$P(A) = P(B_1 \cap \cdots \cap B_{n-1} \cap B)$$

$$= P(B \mid B_1 \cap \cdots \cap B_{n-1}) \cdots P(B_2 \mid B_1)P(B_1)$$

$$= \left(\tfrac{n-1}{N}\right)\left(\tfrac{N-(n-2)}{N}\right)\left(\tfrac{N-(n-3)}{N}\right)\cdots\left(\tfrac{N-1}{N}\right) = \left(\tfrac{n-1}{N}\right)\left(1-\tfrac{1}{N}\right)\left(1-\tfrac{2}{N}\right)\cdots\left(1-\tfrac{n-2}{N}\right)$$

Ejemplo 2.21. *Cada uno de* n *palillos se rompe en una parte corta y una parte larga; inmediatamente después se forman* n *parejas al azar, con las* $2n$ *partes que se obtienen. ¿Cuál es la probabilidad de que a) cada una de las* n *parejas esté formada por una parte corta y una larga? y b) cada una de las* n *parejas quede formada como lo estaba antes de romper los palillos?*

Solución

a. Se puede asumir que la formación de las nuevas n *parejas se forman eligiendo al azar cada una de las parejas de manera consecutiva. Sea entonces* A_k *el evento 'la pareja número* k *está formada por una parte corta y una larga'.*

Dadas r *partes cortas y* r *partes largas, si se seleccionan* 2 *al azar, la probabilidad de que se elija una parte corta y una parte larga está dada por* $\frac{r^2}{\binom{2r}{2}} = \frac{r}{2r-1}$*, así que:*

$$P(A_1 \cap \cdots \cap A_n) = P(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \cdots P(A_2 \mid A_1)P(A_1)$$

$$= \frac{1}{1}\frac{2}{3}\cdots\frac{n-2}{2n-5}\frac{n-1}{2n-3}\frac{n}{2n-1} = \frac{n!(2)(4)\cdots(2n-4)(2n-2)(2n)}{(2n)!} = \frac{2^n(n!)^2}{(2n)!}$$

b. Sea B_k *el evento 'la pareja número* k *forma uno de los palillos originales'. Dados* r *palillos cortados, cada uno en una parte corta y una larga, si se seleccionan* 2 *partes al azar, la probabilidad de que correspondan al mismo palillo está dada por* $\frac{r}{\binom{2r}{2}} = \frac{1}{2r-1}$*, así que:*

$$P(B_1 \cap \cdots \cap B_n) = P(B_n \mid B_1 \cap \cdots \cap B_{n-1}) \cdots P(B_2 \mid B_1)P(B_1)$$

$$= \frac{1}{1}\frac{1}{3}\cdots\frac{1}{2n-5}\frac{1}{2n-3}\frac{1}{2n-1} = \frac{(2)(4)\cdots(2n-4)(2n-2)(2n)}{(2n)!} = \frac{2^n n!}{(2n)!}$$

EJEMPLO 2.22. *En un concurso se colocan tres puertas y detrás de ellas se colocan al azar tres premios, uno de los cuales es un automóvil y los otros dos son pequeños regalos de consolación. A cada concursante se le asigna inicialmente la puerta número* 1*, pero otra persona, que puede ver como están colocados los premios, abre alguna de las puertas* 2 *y* 3 *en donde no se encuentre el automóvil, de tal manera que, si no se encuentra en ninguna de ellas, abre la número* 2 *con probabilidad* p *y la número* 3 *con probabilidad* $1-p$*. El concursante, quien conoce el valor de* p*, tiene entonces la opción de escoger cualquiera de las dos puertas que se encuentren cerradas. ¿Cuál es la mejor estrategia del concursante para tratar de conseguir el automóvil?*

Solución

El espacio muestral del experimento aleatorio está dado por:

$\Omega = \{(1,2),(1,3),(2,3),(3,2)\}$

en donde, en cada pareja de números, el primero indica la puerta detrás de la cual se coloca el automóvil y el segundo indica la puerta que se abre para ser mostrada al concursante.

Utilizando la regla del producto, se tiene:

$P[(1,2)] = \frac{1}{3}p$

$P[(1,3)] = \frac{1}{3}(1-p)$

$P[(2,3)] = \frac{1}{3}$

$P[(3,2)] = \frac{1}{3}$

Definamos ahora los eventos:

A*: el automóvil se encuentra detrás de la puerta número* 1*.*

B*: la puerta abierta que se muestra al concursante es la número* 2*.*

C*: la puerta abierta que se muestra al concursante es la número* 3*.*

El primer análisis que se puede hacer consiste en decir que, como $P(A) = \frac{1}{3}$*, entonces, antes de que se muestre al concursante la puerta que se abre, la mejor estrategia consiste en cambiar de puerta, teniendo así una probabilidad de ganar el automóvil igual a* $\frac{2}{3}$*. Sin embargo conviene hacer un análisis más detallado pues pudiera ser que la ocurrencia de* B *o de* C *incremente la probabilidad de* A*.*

Se tiene $P(A \mid B) = \frac{P(A\cap B)}{P(B)} = \frac{\frac{1}{3}p}{\frac{1}{3}p+\frac{1}{3}} = \frac{p}{p+1}$

Ahora bien, la función $f(p) = \frac{p}{p+1}$*, definida en el intervalo* $[0,1]$*, es creciente, de manera que su valor máximo es* $f(1) = \frac{1}{2}$*. Por lo tanto, si* B *ocurre, la mejor estrategia para el concursante continúa siendo la de cambiar de puerta, independientemente del valor de* p*. Evidentemente, la conclusión es la misma en caso de que* C *ocurra. Así que, en cualquier caso, la mejor estrategia consiste en cambiar de puerta.*

2.7. Independencia estocástica

Hemos visto que la probabilidad condicional de un evento B, dada la ocurrencia de un evento A, es una probabilidad calculada con respecto a un espacio muestral reducido, el que queda conformado únicamente por aquellos posibles resultados del experimento aleatorio para los cuales el evento A ocurre. En general, la probabilidad de ocurrencia de un evento B se altera al reducirse el espacio muestral; tal fue el caso del ejemplo 2.15, en donde se obtiene $P(A_1) = \frac{2}{n(n+1)}$, mientras que al condicionar con la ocurrencia del evento B_m, se tiene $P(A_1|B_m) = \frac{2}{m(m+1)}$. Esto puede interpretarse diciendo que la probabilidad de ocurrencia del evento B no es independiente de la ocurrencia del evento A. De la misma manera, en los casos en que la ocurrencia de un evento A no altere la probabilidad de ocurrencia de un evento B, se puede hablar de independencia de la probabilidad de ocurrencia de B con respecto a la ocurrencia de A. En esta última situación se tiene la siguiente relación:

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

de la cual se sigue, asumiendo $P(B) > 0$:

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B)P(A)}{P(B)} = P(A)$$

Es decir, si la probabilidad de ocurrencia de un evento B es independiente de la ocurrencia de un evento A, entonces la probabilidad de ocurrencia del evento A es también independiente de la ocurrencia de B. En otras palabras, se puede decir simplemente que los eventos A y B son estocásticamente independientes. *El calificativo de estocástico es necesario pues lo que es independiente de la ocurrencia de uno de los eventos es la probabilidad de ocurrencia del otro, y no simplemente su ocurrencia.*

De la condición de independencia estocástica entre dos eventos A y B, se obtiene:

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

La relación entre el primer término y el último de estas igualdades establece una fórmula que tiene la virtud de eliminar la restricción de que las probabilidades de los eventos A y B tengan que ser positivas, como en la forma que utiliza probabilidades condicionales. Por tal motivo, ésta es la forma que se utiliza para definir la independencia estocástica de dos eventos.

DEFINICIÓN 2.23 (**Independencia de 2 eventos**). *Se dice que dos eventos A y B son estocásticamente independientes si $P(A \cap B) = P(A)P(B)$.*

Al igual que la fórmula que establece la definición de probabilidad condicional, la fórmula que establece la independencia estocástica de dos eventos se utiliza en dos sentidos diferentes. Por un lado, una vez que se ha encontrado la probabilidad de cualquier evento, se puede utilizar para determinar cuándo dos eventos dados son estocásticamente independientes. Este uso se ilustra en el ejemplo siguiente:

EJEMPLO 2.24. *En una escuela, las estadísticas muestran que la probabilidad de que un estudiante pase Química es* 0.35*, la probabilidad de que pase Física es* 0.4 *y la probabilidad de que pase ambas es* 0.1*. Sea A el evento 'un estudiante elegido al azar pasa Química' y B el evento 'un estudiante elegido al azar pasa Física'. ¿Son A y B eventos independientes?*

Solución

$P(A)P(B) = (0.35)(0.4) = 0.14 \neq P(A \cap B)$. *Por lo tanto, A y B no son independientes.*

▲

Por otro lado, en ocasiones se sabe, por las características del experimento, que determinadas parejas de eventos resultan independientes y entonces la relación que establece la independencia estocástica de eventos se impone como condición al modelo matemático que se está construyendo. Un ejemplo típico de esta situación se presenta cuando el experimento aleatorio en consideración está compuesto por dos experimentos, los cuales son independientes entre sí y entonces, si A es un evento cuya ocurrencia o no ocurrencia depende exclusivamente de alguno de esos experimentos y B es otro evento cuya ocurrencia o no ocurrencia depende exclusivamente del otro, esos eventos deben de ser estocásticamente independientes dentro de cualquier modelo matemático que se construya para el experimento aleatorio compuesto. Los siguientes 2 ejemplos ilustran estas ideas.

EJEMPLO 2.25. *Una urna contiene* 3 *bolas rojas y* 6 *bolas negras. Un experimento aleatorio consiste en dos partes, en la primera, se selecciona al azar una bola de la urna y se regresa a la misma, en la segunda, nuevamente se selecciona al azar una de las bolas de la urna.*

Denotemos por $B_1, B_2, B_3, B_4, B_5, B_6, B_7, B_8, B_9$ las 9 *bolas de la urna. Cada posible resultado del experimento aleatorio puede entonces representarse mediante una pareja*

(B_i, B_j) con $i, j \in \{1, \ldots, 9\}$, en donde B_i representa el resultado de la primera elección y B_j el de la segunda. El espacio muestral consta así de 81 elementos.

Como la elección de la primera bola se realiza al azar, cualquiera de las bolas en la urna tiene la misma probabilidad de ser elegida, de manera que si, para $i \in \{1, \ldots, 9\}$, definimos los siguientes eventos:

B_i^1: En la primera elección se obtiene la bola B_i,

se tiene, $P(B_i^1) = \frac{1}{9}$ para toda i.

Por otra parte, la elección de la segunda bola se realiza también al azar de entre todas las bolas de la urna original, de manera que si, para $j \in \{1, \ldots, 9\}$, definimos los siguientes eventos:

B_j^2: En la segunda elección se obtiene la bola B_j,

se tiene, $P(B_j^2) = \frac{1}{9}$ para toda j.

Ahora bien, como las dos elecciones son independientes una de la otra, los eventos B_i^1 y B_j^2 ($i, j \in \{1, \ldots, 9\}$) son independientes y entonces se debe de imponer la condición

$P(B_i^1 \cap B_j^2) = P(B_i^1)P(B_j^2) = \frac{1}{81}$

Pero las intersecciones $B_i^1 \cap B_j^2$ representan a los posibles resultados del experimento aleatorio y entonces se concluye que cada uno de ellos tiene una probabilidad igual a $\frac{1}{81}$.

EJEMPLO 2.26. *Un sistema contra incendios está compuesto por dos mecanismos, cada uno de los cuales lo activa automáticamente en caso de un incendio. Se ha determinado, con base en la experiencia, que cuando hay un incendio, el primer mecanismo activa el sistema en 95% de los casos, mientras que el otro mecanismo, que actúa en forma independiente del primero, lo hace en 85% de los casos. Si un incendio comienza en un edificio que tiene instalado dicho sistema, a) ¿cuál es la probabilidad de que el sistema sea activado? b) Dado que el sistema es activado, ¿cuál es la probabilidad de que el primer mecanismo funcione correctamente?*

Solución

Consideremos el evento A: 'el sistema es activado' y, para $i \in \{1, 2\}$, los eventos A_i: 'el mecanismo i activa el sistema'. Se tiene entonces:

a. $P(A) = P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

$= P(A_1) + P(A_2) - P(A_1)P(A_2) = 0.95 + 0.85 - (0.95)(0.85) = 0.9925$

b. $P(A_1 \mid A) = \frac{P(A_1 \cap A)}{P(A)} = \frac{P(A_1)}{P(A)} = \frac{0.95}{0.9925} = 0.957179$

▲

Cabe hacer énfasis en el hecho de que la independencia estocástica entre dos eventos establece una relación entre probabilidades de eventos y no únicamente entre los eventos mismos. Para ilustrar este punto consideremos el ejemplo siguiente:

EJEMPLO 2.27. *Una urna contiene* 2 *bolas rojas y* 2 *bolas negras. Un experimento aleatorio consiste en dos partes, en la primera, se selecciona al azar una bola de la urna y se regresa a la misma, en la segunda, nuevamente se selecciona al azar una de las bolas de la urna. Se considera como resultado del experimento a la pareja ordenada de colores que se obtiene.*

El espacio muestral de este experimento aleatorio se puede representar de la siguiente manera:

$\Omega = \{(B, N), (B, B), (N, B), (N, N)\}$

La probabilidad de cada uno de estos posibles resultados se puede calcular de manera similar a como se obtuvieron las probabilidades de los posibles resultados en el ejemplo previo, obteniéndose que cada uno de ellos tiene una probabilidad igual a $\frac{1}{4}$.

Consideremos ahora los siguientes eventos:

A*: La primera bola seleccionada es roja.*

B*: Las dos bolas seleccionadas son de distinto color.*

Utilizando la propiedad de la aditividad finita, se tiene:

$P(A) = P(\{(B, N), (B, B)\}) = \frac{1}{2}$

$P(B) = P(\{(B, N), (N, B)\}) = \frac{1}{2}$

$P(A \cap B) = P(\{(B, N)\}) = \frac{1}{4}$

Entonces, $P(A \cap B) = P(A)P(B)$*, así que los eventos* A *y* B *son estocásticamente independientes.*

Consideremos ahora el mismo experimento aleatorio, pero con una urna que contiene 2 *bolas rojas y* 4 *bolas negras. El espacio muestral de este nuevo experimento sigue siendo el mismo, a saber:*

$\Omega = \{(R, N), (R, R), (N, R), (N, N)\}$

La probabilidad de cada uno de los posibles resultados se calcula de la misma manera que antes, obteniéndose:

$P(\{(R, N)\}) = \frac{2}{9}$

$P(\{(R, R)\}) = \frac{1}{9}$

$P(\{(N, R)\}) = \frac{2}{9}$

$P(\{(N, N)\}) = \frac{4}{9}$

Considerando los mismos eventos A y B que antes, se tiene:

$P(A) = P(\{(R, N), (R, R)\}) = \frac{3}{9}$

$P(B) = P(\{(R, N), (N, R)\}) = \frac{4}{9}$

$P(A \cap B) = P(\{(R, N)\}) = \frac{2}{9}$

Entonces $P(A \cap B) \neq P(A)P(B)$, así que los eventos A y B no son estocásticamente independientes.

De esta manera hemos mostrado que se puede definir un espacio muestral y dos eventos A y B relativos a ese espacio muestral de tal manera que, con una cierta asignación de probabilidades, los eventos A y B resultan ser estocásticamente independientes, mientras que, con otra asignación de probabilidades, resultan no serlo. La propiedad de independencia involucra entonces a la función de probabilidad de manera determinante.

▲

El concepto de independencia estocástica así como su utilización en los dos sentidos mencionados antes se puede extender al caso de más de dos eventos. Diremos que n eventos $A_1, \ldots, A_n$, relativos todos a un determinado experimento aleatorio, son estocásticamente independientes si la probabilidad de ocurrencia de cualquiera de ellos no es afectada por la ocurrencia de los otros. En otras palabras, si la probabilidad de ocurrencia de cualquier evento A_j ($j \in \{1, \ldots, n\}$) no cambia dada la ocurrencia de cualquier subcolección $A_{j_1}, \ldots, A_{j_m}$, es decir:

$$P(A_j | A_{j_1} \cdots A_{j_m}) = P(A_j),$$

en donde, para $k \in \{1, \ldots, m\}$, $j_k \in \{1, \ldots, n\}$ y $j \notin \{j_1, \ldots, j_m\}$.

Estas relaciones en términos de probabilidades condicionales pueden transformarse en términos de probabilidades de intersecciones, lo cual es más adecuado pues, como en el caso de dos eventos, elimina la restricción de que los eventos en consideración deban tener probabilidad positiva. Es por esto que la definición se da en los siguientes términos:

DEFINICIÓN 2.28 (**Independencia de n eventos**). *Se dice que n eventos $A_1, \ldots, A_n$ son estocásticamente independientes si dada cualquier subcolección de ellos, $A_{j_1}, \ldots, A_{j_m}$, con $1 \leq j_1 < \cdots < j_m$, se tiene:*

$$P(A_{j_1} \cap \cdots \cap A_{j_m}) = P(A_{j_1}) \cdots P(A_{j_m})$$

Debemos observar que, para que n eventos sean estocásticamente independientes, no basta con pedir que lo sean por parejas. Un ejemplo ilustrará este hecho.

EJEMPLO 2.29. *Consideremos el experimento aleatorio consistente en lanzar 2 veces consecutivas una moneda balanceada. El espacio muestral de este experimento se puede representar de la siguiente manera:*

$\Omega = \{(A, S), (A, A), (S, A), (S, S)\}$

en donde S representa cara y A representa cruz.

La probabilidad de cada uno de los elementos del espacio muestral se puede calcular inmediatamente asumiendo que los resultados de ambos lanzamientos son independientes uno del otro, obteniéndose que la probabilidad de cada elemento es igual a $\frac{1}{4}$.

Consideremos ahora los siguientes eventos:

A: Se obtiene cara en el primer lanzamiento.

B: Se obtiene cara en el segundo lanzamiento.

C: Se obtiene el mismo resultado en los dos lanzamientos.

Los eventos A,B y C son estocásticamente independientes por parejas, en efecto, se tiene:

$P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{2}$, $P(A \cap B) = \frac{1}{4}$, $P(A \cap C) = \frac{1}{4}$ *y* $P(B \cap C) = \frac{1}{4}$, *así que,*

$P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$ *y* $P(B \cap C) = P(B)P(C)$.

Pero $P(A \cap B \cap C) = \frac{1}{4}$, *de manera que* $P(A \cap B \cap C) \neq P(A)P(B)P(C)$.

La no independencia de los 3 eventos es intuitivamente clara pues existe una dependencia entre ellos; efectivamente, cuando el evento A ocurre entonces B ocurre si y *sólo si C ocurre. También se puede ver la dependencia entre los 3 eventos observando que dada la ocurrencia de A y B, el evento C se convierte en el evento seguro.*

▲

Por otra parte, también se puede mostrar que dados 3 eventos A, B y C, se puede cumplir la relación $P(A \cap B \cap C) = P(A)P(B)P(C)$ sin que los eventos sean independientes (ver ejercicio 2.27).

EJEMPLO 2.30. *En la búsqueda de un libro, un estudiante tiene acceso a 3 bibliotecas, cada una de las cuales opera en forma independiente de las otras. Para cada biblioteca, la probabilidad de poseer el libro es igual a $\frac{1}{2}$ y, si lo posee, la probabilidad de que se encuentre prestado es también igual a $\frac{1}{2}$. Encuentre la probabilidad de que el estudiante consiga el libro en alguna de las 3 bibliotecas.*

Solución

Consideremos, para $k \in \{1, 2, 3\}$, los eventos A_k: 'la biblioteca k posee el libro', B_k: 'la biblioteca k tiene prestado el libro' y C_k: 'el estudiante consigue el libro en la biblioteca k'. Además, sea C el evento 'el estudiante consigue el libro en alguna de las 3 bibliotecas'. Se tiene entonces:

$$P(C_k^c) = P(A_k \cap B_k) + P(A_k^c) = P(B_k \mid A_k)P(A_k) + P(A_k^c) = \tfrac{1}{2}\tfrac{1}{2} + \tfrac{1}{2} = \tfrac{3}{4}$$

$$P(C) = 1 - P(C^c) = 1 - P(C_1^c \cap C_2^c \cap C_3^c) = 1 - P(C_1^c)P(C_2^c)P(C_3^c)$$

$$= 1 - \left(\tfrac{3}{4}\right)^3 = \tfrac{37}{64} = 0.578125$$

EJEMPLO 2.31. *Calcule la probabilidad de que al lanzar n veces un dado se obtenga por lo menos un 6.*

Solución

Consideremos, para $k \in \{1, \ldots, n\}$, el evento A_k: 'se obtiene 6 en el k-ésimo lanzamiento' y sea B_n el evento 'se obtiene por lo menos un 6 en los n lanzamientos'. Se tiene entonces:

$$P(B_n) = 1 - P\left(B_n^c\right) = P\left(A_1^c \cap \cdots \cap A_n^c\right) = 1 - P\left(A_1^c\right) \cdots P\left(A_n^c\right) = 1 - \left(\tfrac{5}{6}\right)^n$$

Este problema está relacionado con uno de los primeros problemas de probabilidad que se plantearon. Se preguntaba cuántos lanzamientos de un dado se requieren para que sea más favorable obtener por lo menos un 6 que no obtenerlo. Obsérvese que $P(B_n)$ crece con n, siendo su valor más pequeño $\frac{1}{6}$ y su valor límite, cuando $n \rightsquigarrow \infty$, 1. Para alguna n, $P(B_n)$ pasa entonces de un valor menor o igual a $\frac{1}{2}$ a un valor mayor que $\frac{1}{2}$. El problema planteado es entonces equivalente a preguntarse a partir de que n se *tiene* $P(B_n) > \frac{1}{2}$. *Tomando en consideración el valor de* $P(B_n)$ *encontrado arriba, se busca el más pequeño número natural* n *tal que* $\left(\frac{5}{6}\right)^n < \frac{1}{2}$; *es decir* $n \ln \frac{5}{6} < -\ln 2$. *Resolviendo esta desigualdad, se obtiene* $n \geq 4$.

EJEMPLO 2.32. *Cada una de* n *bolas se va colocando al azar en cualquiera de* N *cajas. ¿Cuánto debe valer* n *para que sea más favorable la ocupación de la primera caja?*[2]

Solución

En este caso el experimento consiste en la colocación de n *bolas en* N *cajas, el cual está compuesto de* n *experimentos independientes, consistentes cada uno de ellos en la colocación al azar de una de las bolas en cualquiera de las* N *cajas. Consideremos entonces los eventos siguientes:*

A: *la primera caja es ocupada por alguna de las* n *bolas.*

A_i : *la primera caja es ocupada por la* i*-ésima bola.*

Tenemos entonces:

$$P(A) = 1 - P(A^c) = 1 - P(A_1^c \cap \cdots \cap A_n^c) = 1 - P(A_1^c) \cdots P(A_n^c) = 1 - \left(\frac{N-1}{N}\right)^n$$

El problema planteado se reduce entonces a encontrar un número n *que satisfaga la desigualdad* $\left(\frac{N-1}{N}\right)^n < \frac{1}{2}$, *es decir,* $n > -\frac{\ln 2}{\ln\left(1-\frac{1}{N}\right)}$.

Si N *es grande, se tiene* $\ln\left(1 - \frac{1}{N}\right) \approx -\frac{1}{N}$, *así que entonces se busca* n *tal que* $n\left(-\frac{1}{N}\right) < -\ln 2$, *es decir,* $n > N \ln 2 \approx 0.7N$.

Como aplicación de este resultado, supongamos que en una urna hay 32 *bolas, de las cuales únicamente una está marcada con premio. Si un juego consiste en seleccionar una bola al azar de la urna (reemplazándola), ¿cuántas veces se debe de jugar para que sea más favorable obtener por lo menos un premio?*

$$-\frac{\ln 2}{\ln\left(1-\frac{1}{N}\right)} = -\frac{\ln 2}{\ln\left(1-\frac{1}{32}\right)} = 21.83$$

Así que, jugando por lo menos 22 *veces, la probabilidad de obtener por lo menos un premio es mayor que* $\frac{1}{2}$.

EJEMPLO 2.33. *Supongamos que* n *bolas se van colocando, una por una y al azar, en cualquiera de* n *cajas. Calcule la probabilidad de que la primera caja quede vacía.*

[2] Este problema, el cual incluye como caso particular el problema del ejemplo anterior, fue resuelto por Abraham de Moivre en el año 1717.

Solución

Definamos los eventos:

A_k*: la bola número* k *no se coloca en la primera caja.*

A*: la primera caja queda vacía.*

Se tiene, $P(A) = P(A_1 \cap \cdots \cap A_n) = \left(\frac{n-1}{n}\right)^n = \left(1 - \frac{1}{n}\right)^n$

Si n *es grande, obtenemos la aproximación:*

$P(A) \approx e^{-1} = 0.367879$

Es decir, si n *es grande, la probabilidad de que la primera caja quede vacía es aproximadamente constante e igual a* 0.36788*. En particular, es más probable que la primera caja quede ocupada.*

EJEMPLO 2.34. *Se van colocando al azar bolas, una a una, en cualquiera de* N *cajas, hasta que se ocupe la primera caja, ¿cuál es la probabilidad de que el proceso termine en el paso* n*?*

Solución

Consideremos los siguientes eventos:

A*: el proceso termina en el paso* n*.*

B_{n-1}*: la primera caja no es ocupada por ninguna de las primeras* $n-1$ *bolas.*

B*: la primera caja es ocupada por la* n*-sima bola.*

Se tiene entonces, $P(A) = P(B_{n-1} \cap B) = P(B_{n-1})P(B) = \left(\frac{N-1}{N}\right)^{n-1} \frac{1}{N}$.

2.8. Interpretación objetiva y subjetiva de la probabilidad

Cabe en este momento hacer énfasis en el sentido objetivo que hemos dado a la probabilidad de un evento en los ejemplos considerados hasta este momento. Al construir un modelo para un determinado experimento aleatorio, hemos buscado que la probabilidad de un evento exprese determinadas propiedades objetivas del experimento aleatorio. Así, cuando decimos que la probabilidad de obtener el resultado 5 al lanzar un dado es $\frac{1}{6}$, queremos expresar con eso que las características del experimento consistente en lanzar un dado son tales que el lanzamiento del dado equivale a seleccionar al azar un número entre 1 y 6 y entonces se puede construir el modelo basándonos en la equiprobabilidad de los posibles resultados.

La objetividad en la asignación de probabilidades tiene sentido pues para nosotros un experimento es aleatorio no porque desconozcamos cual resultado se va a obtener, sino porque efectivamente caben varias posibilidades en cuanto al resultado, de acuerdo a como fue planteado en la sección 1.1. En otras palabras, la aleatoriedad del experimento es objetiva.

No es éste el único enfoque que existe en el Cálculo de Probabilidades, también se puede desarrollar éste desde un enfoque subjetivo. Un ejemplo aclarará la diferencia entre dicho enfoque y el que seguimos en este libro.

Ejemplo 2.35. *Supongamos que tenemos dos cajas, una marcada con el número* 1 *y otra marcada con el número* 2*, y que en cada una de ellas hay una bola. Supongamos además que sabemos que una de las bolas es roja y la otra blanca, pero desconocemos cual está en cada caja. Tomamos entonces la caja número* 1 *y nos preguntamos por la probabilidad de encontrar ahí la bola roja.*

Tal y como está planteado el problema, no tenemos definido ahí algún experimento aleatorio pues las bolas ya están colocadas una en cada caja, de manera que al tomar la caja 1 *necesariamente encontraremos en ella la bola que está colocada ahí, habiendo entonces solo un posible resultado. En otras palabras, desde nuestro enfoque, la probabilidad de encontrar la bola roja en la caja* 1 *es* 0 *si en ella se encuentra la bola blanca y es* 1 *si en ella se encuentra la bola roja.*

El azar en este experimento existe sólo en nuestra mente pues si bien la probabilidad buscada es 0 *o* 1*, desconocemos cual de los dos valores es el correcto.*[3]

Ahora bien, con una interpretación subjetiva de la probabilidad, podríamos decir que es igualmente probable encontrar la bola roja o la bola blanca en la caja 1 *y entonces concluir que la probabilidad de encontrar la bola roja en la caja* 1 *es* $\frac{1}{2}$*. Vemos así que los diferentes enfoques nos conducen a resultados diferentes.*

El problema cambia si lo planteamos de la siguiente manera: Supongamos que tenemos dos bolas, una roja y una blanca, y dos cajas, una marcada con el número 1 *y otra marcada con el número* 2*; consideramos entonces el experimento consistente en colocar las bolas al azar, una en cada caja y nos preguntamos por la probabilidad de que la bola roja quede colocada en la caja* 1*.*

Planteado de esta segunda forma, ahora sí tenemos definido un experimento aleatorio. De éste, puede resultar tanto que la bola roja quede colocada en la caja 1*, como que*

[3]Con base en esto se podría definir un experimento consistente en que, una vez seleccionada la caja 1, se elija al azar un color, ya sea rojo o blanco. Tal experimento sería aleatorio y cabría, por ejemplo, preguntarnos por la probabilidad de que el color elegido coincida con el color de la bola de la caja.

quede colocada en la caja 2; y como la colocación es al azar, podemos partir de la equiprobabilidad de los dos posibles resultados. De esta manera la probabilidad de que la bola roja quede colocada en la caja 1 es $\frac{1}{2}$.

Obsérvese que al valor $\frac{1}{2}$ *encontrado de esta manera podemos darle una interpretación frecuentista diciendo que si el experimento aleatorio que hemos definido lo repetimos muchas veces, aproximadamente en la mitad de las veces la bola roja quedará colocada en la caja* 1.

Por otra parte, el valor $\frac{1}{2}$ *encontrado desde un enfoque subjetivo, en la primera forma en que se planteó el experimento, no admite una interpretación frecuentista pues estando ya colocadas las bolas, una en cada caja, una repetición del experimento consiste en colocar las bolas exactamente de la misma manera, así que si el experimento se repite muchas veces, en todas las repeticiones la bola roja quedará en la caja* 1 *o en todas quedará en la caja* 2. *Sin embargo, cabe aclarar que dentro del enfoque subjetivo, este último hecho no constituye una limitación pues dentro de ese enfoque la probabilidad de un evento no es interpretada de acuerdo al principio de regularidad de las frecuencias.*

EJERCICIOS

Ejercicio 2.1. *Un experimento aleatorio admite únicamente dos posibles resultados, uno de ellos ocurre con probabilidad* p *y el otro con probabilidad* p^2. *Encuentre el valor de* p.

Ejercicio 2.2. *Sean* A *y* B *dos eventos tales que* $P(B) = 0.6$ *y* $P(A \cap B) = 0.2$. *Encuentre* $P(A \cup B^c)$.

Ejercicio 2.3. *Sean* A *y* B *dos eventos relativos a un experimento aleatorio y supongamos que la probabilidad de que* A *ocurra es* $\frac{3}{8}$, *la probabilidad de que* B *ocurra es* $\frac{1}{2}$ *y la probabilidad de que ocurra exactamente uno de los dos eventos* A *y* B *es* $\frac{5}{8}$. *Encuentre la probabilidad de que ocurra* A *pero no* B.

Ejercicio 2.4. *Sean* A *y* B *dos eventos relativos a un experimento aleatorio y supongamos que la probabilidad de que* A *ocurra es* $\frac{3}{5}$, *la probabilidad de que* B *ocurra es* $\frac{1}{5}$ *y la probabilidad de que ocurra* A *pero no* B *es* $\frac{1}{2}$. *Encuentre la probabilidad de que a) ocurra al menos uno de los dos eventos* A *y* B *y b) ocurra exactamente uno de los dos eventos* A *y* B.

Ejercicio 2.5. *Sean* A *y* B *dos eventos relativos a un experimento aleatorio. Supongamos que* $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{2}$ *y* $P(A \cap B) = \frac{1}{12}$. *Encuentre a)* $P(A \cup B)$, *b)* $P(A^c \cup B^c)$ *y c)* $P(A^c \cap B)$.

Ejercicio 2.6. *La unión de dos eventos A y B se puede expresar como la unión de dos eventos mutuamente excluyentes, a saber, A y $B - A \cap B$. Exprese la unión de tres eventos A, B y C como unión de eventos mutuamente excluyentes y utilice esa representación para demostrar la siguiente fórmula:*

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Ejercicio 2.7. *Sean A, B y C tres eventos relativos a un experimento aleatorio. Supongamos que $P(A) = \frac{3}{4}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{4}$, $P(A \cap B) = \frac{3}{8}$, $P(A \cap C) = \frac{1}{8}$, $P(B \cap C) = \frac{3}{16}$ y $P(A \cap B \cap C) = \frac{1}{8}$. Encuentre la probabilidad de que no ocurra A, ni B, ni C.*

Ejercicio 2.8 (**Regla de la suma para n eventos**). *Demuestre que si $A_1, \ldots A_n$ son n eventos cualesquiera, entonces:*

$$P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k) - \sum_{\{i,j \in \{1,\ldots,n\}, : i \neq j\}} P(A_i \cap A_j)$$

$$+ \sum_{\{i,j,k \in \{1,\ldots,n\}, : i \neq j, j \neq k, i \neq k\}} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P(A_1 \cap A_2 \ldots \cap A_n)$$

Ejercicio 2.9. *Sean A y B dos eventos con probabilidad positiva tales que $A \subset B$. Demuestre que a) $P(A|B) = P(A)/P(B)$ y b) $P(B|A) = 1$.*

Ejercicio 2.10. *Sean A y B dos eventos de probabilidad positiva tales que $P(A \mid B) < P(A)$, demuestre que $P(B \mid A) < P(B)$.*

Ejercicio 2.11. *Sean A y B eventos con probabilidad positiva tales que $P(A) = P(B)$, demuestre que $P(A|B) = P(B|A)$.*

Ejercicio 2.12. *Un dado desbalanceado está hecho de tal forma que la probabilidad de obtener el número k es igual a ck, en donde c es una constante y $k \in \{1, 2, 3, 4, 5, 6\}$. Un experimento aleatorio consiste en lanzar dicho dado. Dado que se obtiene un número par, ¿cuál es la probabilidad de que se obtenga el número 2?*

Ejercicio 2.13. *En una ciudad se publican los periódicos A, B y C. Una encuesta reciente muestra que 20% de los habitantes adultos de la ciudad lee A, 16% lee B, 14% lee C, 8% lee A y B, 5% lee A y C, 4% lee B y C y 2% lee A, B y C. Si se elige un adulto al azar, calcule la probabilidad de que a) no lea ninguno de los periódicos, b) lea exactamente uno de los periódicos y c) lea al menos A y B sabiendo que lee al menos uno de los 3 periódicos.*

Ejercicio 2.14. *Dada una población formada exclusivamente por hermanos gemelos, consideremos el experimento aleatorio consistente en seleccionar primero una pareja de hermanos gemelos al azar y después, también al azar, uno de los hermanos de esa pareja. Supongamos que la probabilidad de que los hermanos gemelos seleccionados sean del mismo sexo es igual a 0.7 y que la probabilidad de que la persona seleccionada, en la segunda parte del experimento, sea de sexo masculino es igual a 0.4. Sabiendo* *que la persona seleccionada en la segunda parte del experimento es de sexo masculino, ¿cuál es la probabilidad de que su hermano también lo sea?*

Sugerencia: *Represente* Ω *como* $\Omega = \{(MM)M, (FF)F, (MF)M, (MF)F\}$ *y encuentre la probabilidad de cada uno de sus elementos.*

Ejercicio 2.15 (**Regla del producto**). *Demuestre que si* $A_1, \dots A_n$ *son* n *eventos cualesquiera, entonces:*

$$P(\cap_{k=1}^n A_k) = P(A_n|A_1 \cap \ldots \cap A_{n-1}) \cdots P(A_2|A_1)P(A_1)$$

Ejercicio 2.16. *Consideremos una urna que contiene* r *bolas rojas y* b *bolas blancas y un experimento aleatorio que consiste en dos partes, en la primera, se selecciona al azar una bola de la urna y se deja fuera de ésta, en la segunda, se selecciona al azar una de las bolas restantes. Denotemos por* $R_1, \dots, R_r$ *a las bolas rojas. El hecho de que la segunda elección sea al azar se traduce inmediatamente, por ejemplo, en que la probabilidad de obtener en la segunda elección la bola* R_i*, dado que en la primera se obtiene la bola* R_j*, es igual a* $\frac{1}{r+b-1}$*, para cualquier* $i, j \in \{1, \dots, r\}$ *con* $i \neq j$*; por la propiedad de la aditividad finita se sigue entonces que la probabilidad de obtener en la segunda elección una bola roja, dado que en la primera se obtiene la bola* R_j*, es igual a* $\frac{r-1}{r+b-1}$ *para cualquier* $j \in \{1, \dots, r\}$*. Nos gustaría entonces poder decir simplemente que la probabilidad de que la segunda bola seleccionada sea roja dado que la primera también lo es, está dada también por* $\frac{r-1}{r+b-1}$*. Esto es correcto, pero requiere de un pequeño razonamiento que demuestre su validez. Demuestre el siguiente resultado general y utilícelo para fundamentar esta conclusión.*

Sean $A_1, \dots, A_n$ n *eventos mutuamente excluyentes de probabilidad positiva y* B *un evento tal que* $P(B \mid A_1) = \cdots = P(B \mid A_n) = c$*, entonces* $P(B \mid A_1 \cup \cdots \cup A_n) = c$*.*

Ejercicio 2.17. *Las letras A,A,A,C,E,I,M,M,T,T se colocan al azar, una después de la otra. ¿Cuál es la probabilidad de que se obtenga la palabra MATEMATICA?*

Ejercicio 2.18. *Un ropero contiene* 10 *pares (distintos) de zapatos. Si se escogen al azar* 5 *zapatos, ¿cuál es la probabilidad de que la muestra contenga exactamente uno de los pares originales?*

Ejercicio 2.19. *Un ropero contiene* 10 *pares (distintos) de zapatos. Si se escogen al azar* 3 *zapatos, ¿cuál es la probabilidad de que se obtengan* 2 *que formen un par?*

Ejercicio 2.20. *De una caja que contiene* n *pares (distintos) de zapatos se seleccionan al azar* $2r$ *zapatos con* $2r < n$*. Encuentre la probabilidad de que con los zapatos seleccionados a) no se forme algún par y b) se forme exactamente un par.*

Ejercicio 2.21. *En un pueblo de* $n+1$ *habitantes, una persona le rumorea algo a una segunda persona, quien a su vez lo repite a una tercera, etc. En cada caso, la persona escoge al receptor del rumor, al azar, de entre las otras* n *personas del pueblo. Si el* *rumor se transmite r veces ($r < n$), encuentre la probabilidad de que a) no regrese a la que lo originó y b) que no pase dos veces por la misma persona.*

Ejercicio 2.22. *En un concurso se colocan tres puertas y detrás de ellas se colocan tres premios, uno de los cuales es un automóvil y los otros dos son pequeños regalos de consolación. Los premios se colocan de tal manera que la probabilidad de que el automóvil se coloque detrás de la puerta número 1 es igual a q mientras que la probabilidad de que se coloque detrás de la puerta número 2 (resp. 3) es igual a $\frac{1-q}{2}$. A cada concursante se le asigna inicialmente la puerta número 1, pero otra persona, que puede ver como están colocados los premios, abre alguna de las puertas 2 y 3 en donde no se encuentre el automóvil, de tal manera que, si no se encuentra en ninguna de ellas, abre la número 2 con probabilidad p y la número 3 con probabilidad $1 - p$. El concursante, quien conoce los valores de p y q, tiene entonces la opción de escoger cualquiera de las dos puertas que se encuentren cerradas. ¿Cuál es la mejor estrategia del concursante para tratar de conseguir el automóvil?*

Ejercicio 2.23. *Sean A y B dos eventos independientes tales que la probabilidad de que ocurran simultáneamente es $\frac{1}{6}$ y la probabilidad de que ninguno ocurra es $\frac{1}{3}$. Encuentre $P(A)$ y $P(B)$.*

Ejercicio 2.24. *Sean A y B dos eventos independientes y supongamos $P(A) = \frac{1}{2}$ y $P(A \cup B) = \frac{7}{10}$. Encuentre $P(B)$.*

Ejercicio 2.25. *Sean A y B dos eventos relativos a un experimento aleatorio. Supongamos que $P(A) = 0.4$, $P(A \cup B) = 0.7$ y sea $p = P(B)$. a) ¿Para qué valor de p son A y B mutuamente excluyentes? b) ¿Para que valor de p son A y B independientes?*

Ejercicio 2.26. *Un experimento aleatorio consiste en lanzar un par de dados, uno blanco y uno azul. Sea A el evento 'se obtiene 1 o 2 con el dado blanco' y sea B el evento 'la suma de los números que se obtienen es igual a 7'. ¿Son A y B eventos independientes? Justifique su respuesta.*

Ejercicio 2.27. *Muestre con un ejemplo que es posible tener 3 eventos A, B y C tales que $P(A \cap B \cap C) = P(A)P(B)P(C)$ pero de tal manera que estos eventos no sean independientes.*

Ejercicio 2.28. *Se tienen dos dados, uno rojo y uno azul. Un experimento aleatorio consiste en lanzar el par de dados dos veces consecutivas. Encuentre la probabilidad de que se obtenga exactamente el mismo resultado en los dos lanzamientos, es decir, el mismo número con el dado rojo y el mismo número con el dado azul.*

Ejercicio 2.29. *Calcule el número de lanzamientos de un par de dados que se requieren para que sea más favorable obtener por lo menos un par de seises que no obtenerlo.*[4]

[4]Este problema fue resuelto en forma correcta en el año 1654 por Blaise Pascal, siendo uno de los primeros problemas de probabilidad relativamente complejo que se planteó.

EJERCICIO 2.30. *Calcule el número de lanzamientos de tres dados que se requieren para que sea más favorable obtener por lo menos una tercia de cincos que no obtenerla.*

EJERCICIO 2.31. *Cada una de n urnas contiene N bolas numeradas de 1 a N. Se extrae una bola al azar de cada urna. Calcule la probabilidad de que m sea el más grande número extraído, en donde $m \in \{1, \ldots, N\}$.*

EJERCICIO 2.32. *Un experimento aleatorio consiste en lanzar sucesivamente una moneda balanceada hasta que se obtengan dos resultados iguales en forma consecutiva. Encuentre la probabilidad de que el experimento termine en el sexto lanzamiento.*

EJERCICIO 2.33. *Una urna contiene 2 bolas negras y 4 bolas blancas. Un experimento aleatorio consiste en ir seleccionando al azar una bola de la urna, regresándola después de observar su color, hasta que salgan dos bolas del mismo color en forma consecutiva. Encuentre la probabilidad de que este proceso termine antes de la cuarta elección.*

EJERCICIO 2.34. *Una máquina está compuesta por 4 componentes que funcionan independientemente uno del otro y que están organizados de tal manera que la máquina falla únicamente si los 4 componentes fallan. Supongamos que las probabilidades de que cada componente falle son 0.1, 0.2, 0.25 y 0.3, respectivamente. ¿Cuál es la probabilidad de que la máquina funcione bien?*

EJERCICIO 2.35. *Un componente de una cierta máquina falla el 20% de las veces. Con el objeto de disminuir la probabilidad de que la máquina falle, se instalan en la máquina n de dichos componentes de tal manera que la máquina falla únicamente cuando los n componentes fallan. Asumiendo que los n componentes funcionan de manera independiente, ¿cuál es el más pequeño valor de n que garantiza que la máquina funcionará bien el 99% de las veces?*

EJERCICIO 2.36. *Dos equipos A y B van a jugar una serie de juegos de beisbol. El equipo que gane 2 de 3 juegos gana la serie. El primer juego se realizará en el estadio del equipo A, el segundo se hará en el estadio del equipo B y, en caso de llegar a un tercer juego, el último juego se efectuará en el estadio del equipo B. Se sabe que cuando juega en su estadio, el equipo A tiene una probabilidad de ganarle al equipo B igual a 0.7, mientras que cuando se juega en el estadio del equipo B, la probabilidad de que el equipo A le gane al equipo B es igual a 0.2. Suponiendo que los resultados de los juegos son independientes entre sí, calcule la probabilidad de que el equipo A gane la serie.*

EJERCICIO 2.37. *En una cierta compañía el esquema para aprobar una propuesta es el siguiente: tres personas —A, B y C— analizan la propuesta y ésta es aprobada únicamente si por lo menos dos de las tres personas dan su visto bueno. Se sabe que las tres personas analizan cada propuesta en forma independiente y que las probabilidades de que den su visto bueno ante una propuesta son 0.3, 0.2 y 0.1, respectivamente. a) ¿Cuál es la probabilidad de que una determinada propuesta sea aprobada. b) Si se* *sabe que una propuesta es aprobada, ¿cuál es la probabilidad de que ésta sea aprobada por C?*

Ejercicio 2.38. *En una cierta compañía, la toma de decisiones sigue el esquema que se muestra a continuación: Cualquier propuesta pasa primero por A; si la aprueba, entonces la propuesta se pasa a B, C y D; si ya sea B o D la aprueban, la propuesta pasa entonces a E; si ya sea E o C aprueban la propuesta entonces se pasa a F. La propuesta es aprobada finalmente únicamente si ésta llega hasta F y a su vez F la aprueba. Supongamos que la probabilidad de que A, C y F aprueben cualquier propuesta que les llegue es* 0.5*, mientras que la probabilidad de que B, D y E aprueben cualquier propuesta que les llegue es* 0.7*. Supongamos además que A, B, C, D, E y F toman sus decisiones independientemente uno del otro. a) ¿Cuál es la probabilidad de que una determinada propuesta sea aprobada? b) Si se sabe que una propuesta es aprobada, ¿cuál es la probabilidad de que pase por E? c) Si se sabe que una propuesta es aprobada, ¿cuál es la probabilidad de que pase por C y éste la apruebe?*

Ejercicio 2.39. *Dos personas P y Q juegan a lanzar consecutivamente una moneda balanceada con la condición de que cada vez que se obtenga cara, P gana un punto, mientras que Q lo gana cuando se obtiene cruz. Cada persona apuesta una cantidad x y convienen en que gana el juego quien obtenga primero* 4 *puntos. Supongamos que, cuando el jugador P lleva ganado un punto y el jugador Q dos puntos, se les pierde la moneda y no pueden continuar el juego. ¿Cómo deben repartirse las apuestas en ese momento de manera que se tome en cuenta correctamente el número de puntos que lleva ganado cada uno?*

Ejercicio 2.40. *Sean A y B dos eventos independientes. Demuestre que también son independientes a) A^c y B^c, b) A y B^c y c) A^c y B.*

Ejercicio 2.41. *Sean A, B y C tres eventos relativos a un experimento aleatorio. ¿Cuáles de las siguientes aseveraciones son verdaderas?*

a) $P(B|A) + P(B^c|A) = 1$

b) $P(B|A) + P(B|A^c) = 1$

c) Si A y B son independientes, entonces $P(A \cap B|C) = P(A|C)P(B|C)$.

d) Si $P(A|B) = P(B)$*, entonces A y B son independientes.*

e) Si $P(A) = P(B)$*, entonces* $P(A|B) = P(B|A)$.

f) Si $P(A|B) = P(B|A)$*, entonces* $P(A) = P(B)$.

g) Si $P(A) = P(B) = p$*, entonces* $P(A \cap B) \leq p^2$.

En todos los casos, justifique su respuesta, ya sea demostrando que la aseveración es verdadera o dando un contraejemplo que muestre que la aseveración es falsa.

Ejercicio 2.42. *Dos de tres prisioneros son elegidos al azar para ser liberados. El prisionero A pide al guardia que investigue los nombres seleccionados y que le diga uno de ellos que no sea él mismo. Supongamos que el guardia acepta hacer lo que le pide el prisionero y que, en caso de que A no vaya a ser liberado, le dirá el nombre del prisionero B con probabilidad p y el del prisionero C con probabilidad $1-p$. Consideremos entonces los siguientes eventos:*

A: El prisionero A es seleccionado para ser liberado.

B: El guardia informa al prisionero A que el prisionero B va a ser liberado.

¿Son A y B independientes?

Ejercicio 2.43. *Consideremos la situación del ejercicio 2.22 y definamos los siguientes eventos:*

A: El automóvil se encuentra detrás de la puerta número 1.

B: La puerta abierta que se muestra al concursante es la número 2.

¿Son A y B independientes?

CAPÍTULO 3

MUESTREO ALEATORIO

El arte de hacer Matemáticas consiste en encontrar ese caso particular que contiene todos los gérmenes de generalidad.

David Hilbert

3.1. Muestreo aleatorio con reemplazo

Supongamos que tenemos una población de N objetos distintos. Siendo distintos, podemos pensar en que los N objetos están marcados de tal manera que se pueden distinguir uno del otro[1]. Consideremos entonces el siguiente experimento: elegimos un objeto al azar de entre los N, anotamos el resultado y regresamos el objeto seleccionado a la población; elegimos nuevamente un objeto al azar de entre los N, anotamos el resultado y regresamos el objeto seleccionado a la población; este proceso lo continuamos hasta haber realizado n elecciones. A un experimento aleatorio de este tipo lo llamaremos un **muestreo aleatorio con reemplazo**, a los n datos obtenidos los llamaremos una **muestra con reemplazo de tamaño** n y al número n lo llamaremos el **tamaño de la muestra**.

El resultado de un muestreo aleatorio con reemplazo consiste de una colección de n datos, los cuales representan los n objetos que se obtienen en las n elecciones. Si denotamos a los N objetos de la población por $y_1, ..., y_N$, entonces cada posible resultado lo podemos representar por una eneada $(y_{k_1}, \ldots, y_{k_n})$, en donde cada índice k_j es un número entero entre 1 y N y y_{k_j} es el elemento que se obtiene en la j-ésima elección. El espacio muestral consta así de N^n elementos.

[1] En un problema real, los N objetos pueden ser indistinguibles para un observador. El pensarlos como distinguibles es sólo para fines de entendimiento del modelo y se justifica por el hecho de que en cualquier caso, aunque pudiendo ser indistinguibles, se trata de N objetos distintos.

Como la elección de cada una de los objetos se realiza al azar, en cada elección, cualquiera de las objetos de la población tiene la misma probabilidad de ser elegido, de manera que si, para $i \in \{1, \dots, N\}$ y $j \in \{1, \dots, n\}$, definimos los siguientes eventos:

B_i^j: en la j-ésima elección se obtiene el elemento y_i.

se tiene, $P(B_i^j) = \frac{1}{N}$ para toda i.

Ahora bien, como las n elecciones son independientes entre sí, los eventos de cualquier colección $B_{k_j}^j$ $(j \in \{1, \dots, n\})$ son independientes y entonces se debe de imponer la condición:

$P(B_{k_1}^1 \cap \cdots \cap B_{k_n}^n) = P(B_{k_1}^1) \cdot P(B_{k_n}^n) = \frac{1}{N^n}$

Pero las intersecciones $B_{k_1}^1 \cap \cdots \cap B_{k_n}^n$ representan a los posibles resultados del experimento aleatorio y entonces se concluye que cada uno de ellos tiene una probabilidad igual a $\frac{1}{N^n}$.

Conclusión: En un muestreo aleatorio con reemplazo, los N^n posibles resultados son equiprobables.

Obsérvese que el experimento aleatorio consistente en lanzar n dados, el cual es equivalente a lanzar n veces el mismo dado, se puede pensar como un muestreo aleatorio con reemplazo. La población consiste de las 6 caras del dado y la muestra que se obtiene es de tamaño n.

Ejemplo 3.1. *Un experimento aleatorio consiste en lanzar un par de dados, ¿cuál es la probabilidad de que la suma de los números que se obtienen sea igual a* 9*?*

Solución

Tratándose de un muestreo aleatorio con reemplazo, cada uno de los 36 *posibles resultados del experimento tiene una probabilidad igual a* $\frac{1}{36}$. *Ahora bien, de los* 36 *posibles resultados, hay* 4 *en los cuales la suma es igual a* 9*, por lo tanto la probabilidad buscada es igual a* $\frac{1}{9}$.

Obsérvese que, como estamos interesados únicamente en la suma de los números que se obtienen, independientemente del resultado que se obtenga con cada dado, podríamos pensar los posibles resultados del experimento como las diferentes sumas que se pueden obtener. En ese caso el espacio muestral tendría únicamente 11 *elementos, a saber cualquier número entero entre* 2 *y* 12*. Sin embargo, esto no simplifica el problema pues los elementos de ese espacio muestral no son equiprobables.*

También podríamos pensar los posibles resultados del experimento como parejas, de tal manera que una pareja (a, b) *representara el resultado consistente en la obtención de* a *con un dado y* b *con el otro, sin importar cual resultado específico se obtiene con cada dado. En este caso, el espacio muestral constaría de* 21 *posibles resultados, los cuales se describen en la siguiente figura:*

$$
\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
 & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
 & & (3,3) & (3,4) & (3,5) & (3,6) \\
 & & & (4,4) & (4,5) & (4,6) \\
 & & & & (5,5) & (5,6) \\
 & & & & & (6,6)
\end{array}
$$

Sin embargo, este método tampoco simplificaría el problema pues nuevamente los 21 *elementos del espacio muestral no son equiprobables.*

El experimento consistente en lanzar dos dados es equivalente al de lanzar 2 *veces el mismo dado y entonces no se trata ya de* 2 *dados distintos. Sin embargo, en este caso lo que habría que distinguir no son los dos dados sino los lanzamientos pues se trata de dos lanzamientos distintos.*

3.2. Muestreo aleatorio ordenado sin reemplazo

Supongamos nuevamente que tenemos una colección de N objetos distintos y consideremos el siguiente experimento aleatorio: elegimos un objeto al azar de entre los N; después, sin regresar el primer objeto a la población, elegimos al azar otro objeto de entre los que restan y así sucesivamente, hasta tener elegidos n objetos. A un experimento aleatorio de este tipo lo llamaremos un **muestreo aleatorio ordenado sin reemplazo**, a los n datos obtenidos lo llamaremos una **muestra ordenada sin reemplazo de tamaño** n y al número n lo llamaremos el **tamaño de la muestra**.

Como en el caso anterior, el resultado de un muestreo aleatorio sin reemplazo consiste de una colección de n datos, los cuales representan los n objetos que se obtienen en las n elecciones. Si denotamos a los N objetos de la población por $y_1, \ldots, y_N$, entonces cada posible resultado lo podemos representar por una eneada $(y_{k_1}, \ldots, y_{k_n})$, en donde cada índice k_j es un número entero entre 1 y N y todos ellos son distintos.

Como las elecciones de los objetos se realizan en orden, los posibles resultados son eneadas ordenadas $(y_{k1}, \ldots, y_{kn})$, en donde y_{kj} representa el resultado de la j-ésima elección; así, dos eneadas que tengan los mismos elementos pero en distinto orden, representan dos resultados distintos. El espacio muestral consta así de $N(N-1)\cdots(N-n+1)$ elementos.

Denotaremos por $(N)_n$ al producto $N(N-1)\cdots(N-n+1)$ y denominaremos a esta cantidad el **número de ordenaciones de N objetos tomados de n en n.** Es evidente que se tiene la siguiente relación:

$$(N)_n = \frac{N!}{(N-n)!}$$

Como la elección de cada uno de los objetos se realiza al azar, en cada elección cualquiera de las objetos que quedan de la población tiene la misma probabilidad de ser elegido, de manera que si, para $i \in \{1,\dots,n\}$ y $j \in \{1,\dots,N\}$, definimos los siguientes eventos:

B_i^j: En la j-ésima elección se obtiene el elemento y_i,

se tiene, $P(B_i^1) = \frac{1}{N}$ para $i \in \{1,\dots,N\}$. Además, $P(B_{k_j}^j | B_{k_1}^1 \cap \cdots \cap B_{k_{j-1}}^{j-1}) = \frac{1}{N-j+1}$ para $j \in \{2,\dots,N\}$, en donde cada índice k_i es un número entero entre 1 y N y todos ellos son distintos.

Entonces utilizando la regla del producto, se obtiene:

$P(B_{k_1}^1 \cap \cdots \cap B_{k_n}^n)$

$= P(B_{k_n}^n | B_{k_1}^1 \cap \cdots \cap B_{k_{n-1}}^{n-1}) P(B_{k_{n-1}}^{n-1} | B_{k_1}^1 \cap \cdots \cap B_{k_{n-2}}^{n-2}) \cdots$

$P(B_{k_3}^3 | B_{k_1}^1 \cap B_{k_2}^2) P(B_{k_2}^2 | B_{k_1}^1) P(B_{k_1}^1)$

$= \frac{1}{N-n+1} \frac{1}{N-n+2} \cdots \frac{1}{N-2} \frac{1}{N-1} \frac{1}{N} = \frac{1}{N(N-1)(N-2)\dots(N-n+1)}$

Pero las intersecciones $B_{k_1}^1 \cap \cdots \cap B_{k_n}^n$ representan a los posibles resultados del experimento aleatorio y entonces se concluye que cada uno de ellos tiene una probabilidad igual a:

$\frac{1}{N(N-1)(N-2)\dots(N-n+1)}$

Conclusión: En un muestreo aleatorio ordenado sin reemplazo, los $N(N-1)\cdots(N-n+1)$ posibles resultados son equiprobables.

Ejemplo 3.2. *Una urna contiene r bolas rojas y b bolas blancas. Se elige al azar una bola de la urna, después, sin reemplazar la primera, se elige otra y así sucesivamente hasta sacar todas las bolas. Calcule la probabilidad de que en el j-ésimo paso se saque una bola roja.*

Solución

Podemos suponer que el proceso se detiene en el j-ésimo paso y entonces estamos tomando una muestra aleatoria ordenada sin reemplazo de tamaño j, de una población de tamaño $r+b$; así que, tomando muestras ordenadas, hay un total de $(r+b)_j$ posibles resultados equiprobables. De éstos, aquellos en los cuales se obtiene una bola roja en el último paso son $(r+b-1)_{j-1} \cdot r$ pues una bola roja puede ser cualquiera de las r y una vez definida ésta quedan $r+b-1$ bolas, de las cuales hay que elegir $j-1$. Entonces, llamando R_j al evento en consideración, se obtiene:

$P(R_j) = \frac{(r+b-1)_{j-1} \cdot r}{(r+b)_j} = \frac{(r+b-1)(r+b-2)\cdots(r+b-j+1)r}{(r+b)(r+b-1)\cdots(r+b-j+1)} = \frac{r}{r+b}$

También se puede hacer el razonamiento de la siguiente manera:

Sean $a_1, \ldots, a_r$ las bolas rojas y $a_{r+1}, \ldots, a_{r+b}$ las bolas blancas. Denotemos entonces por s_k al total de muestras ordenadas sin reemplazo de tamaño $r+b$ en las cuales la bola k se elige en el j-ésimo paso y por s al total de muestras ordenadas sin reemplazo de tamaño $r+b$. Se tiene entonces $s_1 = s_2 = \cdots = s_{r+b}$ y $s = s_1 + s_2 + \cdots + s_{r+b}$, así que:

$P(R_j) = \frac{s_1+s_2+\cdots+s_r}{s} = \frac{rs_1}{(r+b)s_1} = \frac{r}{r+b}$

3.3. Muestreo aleatorio no ordenado sin reemplazo

Supongamos nuevamente que tenemos una colección de N objetos distintos y consideremos el experimento aleatorio consistente en elegir al azar un paquete de n objetos de entre los N. A un experimento aleatorio de este tipo lo llamaremos un **muestreo aleatorio no ordenado sin reemplazo**, a los n datos obtenidos lo llamaremos una **muestra no ordenada sin reemplazo de tamaño** n y al número n lo llamaremos el **tamaño de la muestra**.

Como en el caso anterior, el resultado de un muestreo aleatorio no ordenado sin reemplazo consiste de una colección de n datos, los cuales representan los n objetos que se seleccionan. Si denotamos a los N objetos de la población por $y_1, \ldots, y_N$, entonces como la elección de los objetos se realiza en paquete, cada posible resultado es una eneada no ordenada, es decir,un subconjunto de tamaño n del conjunto $\{y_1, \ldots, y_N\}$. Así, cada posible resultado puede representarse por $\{y_{k_1}, \ldots, y_{k_n}\}$, en donde cada índice k_j es un número entero y $1 \leq y_{k_1} < \cdots < y_{k_n} \leq N$.

Como la elección de cada uno de los paquetes se realiza al azar, cada uno de los subconjuntos $\{y_{k_1}, \ldots, y_{k_n}\}$ tiene la misma probabilidad de ser seleccionado, de manera que se concluye que también en un muestreo aleatorio no ordenado sin reemplazo, los posibles resultados son equiprobables.

Siendo cada posible resultado del experimento un subconjunto de tamaño n del conjunto $\{y_1, \ldots, y_N\}$, el número de elementos del espacio muestral es igual al número de subconjuntos de tamaño n que se pueden obtener de un conjunto de N elementos. Es bien sabido como calcular este número; aquí lo podemos obtener argumentando que por cada muestra no ordenada sin reemplazo de tamaño n, permutando sus elementos, se obtienen $n!$ muestras ordenadas sin reemplazo de tamaño n, así que el total de muestras no ordenadas sin reemplazo de tamaño n es igual a $\frac{(N)_n}{n!} = \frac{N!}{n!(N-n)!}$, cantidad que es conocida como el número de combinaciones de N objetos tomados de n en n, y que se representa con la notación $\binom{N}{n}$.

Este razonamiento muestra de paso una propiedad importante, a saber, que cuando se tiene un muestreo no ordenado sin reemplazo, la probabilidad de un evento puede calcularse también asumiendo que el muestreo es ordenado. Esto se debe a que, como se dice arriba, por cada muestra no ordenada sin reemplazo, se obtienen $n!$ muestras ordenadas, las cuales corresponden a un muestreo ordenado sin reemplazo y, como tanto las muestras ordenadas como las no ordenadas son equiprobables, el evento formado por las $n!$ muestras ordenadas que corresponden a una muestra no ordenada tiene la misma probabilidad de ocurrencia que tal muestra no ordenada. De esta manera, cuando el experimento aleatorio consiste en un muestreo sin reemplazo y se busca encontrar la probabilidad de un evento en el cual no está involucrado el orden de los elementos de la muestra, no es necesario especificar si el muestreo se realiza con orden o sin orden pues se obtiene la misma solución asumiendo cualquiera de los dos casos. Obviamente, cuando en el evento cuya probabilidad nos interese calcular está involucrado el orden de los elementos de la muestra, necesariamente tenemos que asumir que el muestreo es ordenado y debemos de tomar como posibles resultados a las posibles muestras ordenadas que puedan obtenerse.

En resumen, hemos visto que en los 3 casos que hemos considerado de muestreo aleatorio, los posibles resultados del experimento son equiprobables, de tal manera que, de acuerdo con la propiedad de la aditividad finita, para calcular la probabilidad de un evento cualquiera, relativo al muestreo aleatorio en consideración, basta con determinar el número de posibles resultados que lo componen y la probabilidad buscada es entonces el cociente de ese número entre el total de posibles resultados.

El siguiente resultado se demuestra fácilmente:

Proposición 3.3. *Sean $n, N \in \mathbb{N}$ con $n < N$, entonces:*

$\binom{N}{n} = \binom{N}{N-n}$

$\binom{N}{n} + \binom{N}{n+1} = \binom{N+1}{n+1}$

Ejemplo 3.4. *Supongamos que n bolas se van colocando una por una, al azar, en cualquiera de n cajas. Calcule la probabilidad de que sólo la caja 1 quede vacía.*

Solución

Consideremos los siguientes eventos:

A: únicamente la caja 1 queda vacía.

A_i: únicamente la caja 1 queda vacía y quedan dos bolas en la caja i ($i \in \{2, \ldots, n\}$).

La colocación de las n bolas en las n cajas puede interpretarse como un muestreo con reemplazo de tamaño n, de una población de tamaño n, de manera que se tiene:

$P(A_i) = \frac{\binom{n}{2}(n-2)!}{n^n}$

Y entonces, $P(A) = P(A_2) + \cdots + P(A_n) = \frac{(n-1)\binom{n}{2}(n-2)!}{n^n} = \frac{\binom{n}{2}(n-1)!}{n^n}$

Ejemplo 3.5. *En una urna hay N tarjetas numeradas del 1 al N. Se toma una muestra, al azar y sin reemplazo, de tamaño n. Calcule la probabilidad de que el número más grande de la muestra sea igual a k, en donde $k \geq n$.*

Solución

El problema se puede resolver de dos maneras distintas:

1o. Asumiendo que el muestreo es no ordenado, tenemos un total de $\binom{N}{n}$ posibles resultados y éstos son equiprobables. Para que un posible resultado pertenezca al evento en consideración se requiere que de la muestra de n elementos, $n-1$ sean números entre 1 y $k-1$ y uno de ellos sea igual a k; es decir, $n-1$ de ellos se deben elegir como una muestra, no ordenada, de una población de $k-1$ elementos, de las cuales hay un total de $\binom{k-1}{n-1}$; el objeto restante se puede elegir de una sola manera. Entonces, llamando A al evento en consideración, tenemos:

$P(A) = \frac{\binom{k-1}{n-1}}{\binom{N}{n}}$

2o. Definamos el evento A_k como 'el número más grande de la muestra es menor o igual a k', entonces $A = A_k - A_{k-1}$, de manera que, nuevamente asumiendo que el muestreo es no ordenado:

$P(A) = P(A_k) - P(A_{k-1}) = \frac{\binom{k}{n}}{\binom{N}{n}} - \frac{\binom{k-1}{n}}{\binom{N}{n}} = \frac{\binom{k-1}{n-1}}{\binom{N}{n}}$

EJEMPLO 3.6. *De una población de N objetos se toma una muestra aleatoria con reemplazo de tamaño n. Calcule la probabilidad de que todos los elementos de la muestra sean distintos.*

Solución

En este caso tenemos un total de N^n posibles resultados, los cuales son equiprobables. De éstos, el número de casos en que todos los elementos de la muestra con distintos es igual al número total de muestras ordenadas sin reemplazo, es decir $(N)_n$. Entonces llamando A al evento en consideración, tenemos:

$$P(A) = \frac{(N)_n}{N^n} = \frac{\binom{N}{n}n!}{N^n} = \frac{N\cdots(N-n+1)}{N^n} = (1-\frac{1}{N})\cdots(1-\frac{n-1}{N})$$

Obsérvese que cuando el tamaño de la población es muy grande, comparado con el tamaño de la muestra, la probabilidad anterior es muy cercana a 1. En otras palabras, cuando el tamaño de la población es muy grande, el total de resultados equiprobables en un muestreo con reemplazo es del mismo orden de magnitud que el total de resultados equiprobables en un muestreo sin reemplazo. Esto es intuitivamente evidente pues si la población es grande, es de esperarse que ningún elemento se obtenga dos veces o más. Este resultado puede interpretarse en el sentido de que para poblaciones grandes el muestreo con reemplazo y el muestreo sin reemplazo son prácticamente idénticos.

Este ejemplo se presenta en una diversidad de situaciones; como las siguientes:

a) Un número de 4 cifras se forma eligiendo dígitos al azar del conjunto $\{0, 1, ..., 9\}$. ¿Cuál es la probabilidad de que los 4 dígitos sean distintos? En este caso, $N = 10$ y $n = 4$, así que:

$$P(A) = \frac{\binom{10}{4}4!}{10^4} = \frac{63}{125} = 0.504$$

en donde A es el evento 'los 4 dígitos seleccionados son distintos'.

b) Supongamos que elegir una persona al azar y anotar el día de su cumpleaños es equivalente a elegir al azar un día de entre los 365 del año. ¿Cuál es entonces la probabilidad de que en un grupo de 23 personas, elegidas al azar, todas tengan su cumpleaños en días distintos?

En este caso, $N = 365$ y $n = 23$, así que:

$$P(A) = \frac{\binom{365}{23}(23)!}{(365)^{23}} = 0.4927$$

en donde A es el evento 'todas las personas seleccionadas tienen su día de cumpleaños en días distintos'.

Esta probabilidad es menor que $\frac{1}{2}$; si el grupo fuera de 22 personas obtendríamos una probabilidad mayor que $\frac{1}{2}$ (0.5243). Entonces, en una población de 22 personas es más probable que todas festejen su cumpleaños en días distintos, pero, en una población de 23 o más, es más probable que coincida el día de cumpleaños de por lo menos dos personas.

c) Un elevador parte con 7 pasajeros y se detiene en 10 pisos. Si suponemos que cada persona desciende del elevador, al azar, en cualquiera de los pisos, calcule la probabilidad de que todas las personas bajen del elevador en pisos distintos.

En este caso $N = 10$ y $n = 7$, así que:

$$P(A) = \frac{\binom{10}{7}7!}{10^7} = 0.06048$$

en donde A es el evento 'las 7 personas bajan del elevador en pisos distintos'.

EJEMPLO 3.7. *n bolas se encuentran colocadas en n cajas, una en cada caja. Un experimento consiste en sacar las bolas de la caja y en colocarlas después, al azar, nuevamente una en cada caja. Calcule la probabilidad de que, al realizar el experimento, ninguna bola quede en la caja en la que estaba originalmente.*

Solución

El lector puede intentar calcular esta probabilidad directamente y se dará cuenta que no resulta tan simple. El uso de la regla de la suma nos facilitará este cálculo.

Definamos los siguientes eventos:

A: ninguna bola queda en la caja en la que estaba originalmente.

B: por lo menos una bola queda en la caja en la que esta originalmente.

B_i : la bola que estaba colocada en la i-ésima caja queda en esa misma caja.

Se tiene $B = B_1 \cup \cdots \cup B_n$, entonces usando la regla de la suma, tenemos:

$$P(B) = P(B_1) + \cdots + P(B_n) - P(B_1 \cap B_2) - P(B_1 \cap B_3) - \cdots - P(B_{n-1} \cap B_n) + \cdots$$

Para encontrar $P(B)$, basta entonces con calcular $P(B_{i_1} \cap \cdots \cap B_{i_r})$ para cualquier colección de índices $1 \leq i_1 < i_2 < \cdots < i_r \leq n$. Pero $B_{i_1} \cap \cdots \cap B_{i_r}$ es el evento consistente en que las bolas que estaban colocadas en las cajas $i_1, \ldots, i_r$ quedan en su lugar.

Como nos interesa distinguir el lugar que ocupa cada bola, el experimento aleatorio es equivalente a un muestreo ordenado sin reemplazo en donde tanto el tamaño de *la población como el de la muestra es n. Tenemos entonces un total de $n!$ posibles resultados, los cuales son equiprobables. De éstos, aquellos en los cuales las bolas que estaban en las cajas $i_1, \dots, i_r$ quedan en su lugar, se obtienen permutando los lugares de las $n-r$ bolas restantes; es decir, obtenemos $(n-r)!$ resultados favorables al evento $B_{i_1} \cap \cdots \cap B_{i_r}$. Entonces:*

$P(B_{i_1} \cap \cdots \cap B_{i_r}) = \frac{(n-r)!}{n!}$

Notemos que esta probabilidad depende sólo del número de índices y no de los índices específicos de que se trata, así que, como en la expresión que se obtiene para $P(B)$ hay $\binom{n}{r}$ sumandos de la forma $P(B_{i_1} \cap \cdots \cap B_{i_r})$, utilizando la regla de la suma, se tiene:

$P(B) = \binom{n}{1}\frac{(n-1)!}{n!} - \binom{n}{2}\frac{(n-2)!}{n!} + \binom{n}{3}\frac{(n-3)!}{n!} - \cdots + (-1)^{n+1}\binom{n}{n}\frac{1}{n!}$

$= \frac{n!}{1!(n-1)!}\frac{(n-1)!}{n!} - \frac{n!}{2!(n-2)!}\frac{(n-2)!}{n!} + \frac{n!}{3!(n-3)!}\frac{(n-3)!}{n!} - \cdots + (-1)^{n+1}\frac{1}{n!}$

$= \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1}\frac{1}{n!}$

y entonces:

$P(A) = 1 - P(B) = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n}\frac{1}{n!}$

es la probabilidad buscada.

Podemos encontrar una aproximación de esta probabilidad si recordamos la relación $e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots$, de la cual se obtiene $e^{-1} = \frac{1}{2!} - \frac{1}{3!} + \cdots$. Entonces, si n es grande, obtenemos:

$P(A) \approx e^{-1} = 0.367879$

Es decir, si n es grande, la probabilidad de que ninguna bola quede en la caja en la que estaba colocada originalmente es aproximadamente constante e igual a 0.368. Obsérvese entonces, en particular, que, con probabilidad mayor que $\frac{1}{2}$, por lo menos una bola queda colocada en la caja donde estaba.

Para $n = 6$, el resultado exacto es $P(A) = \frac{53}{144} = 0.368056$, así que a partir de esta n, podemos considerar como buena la aproximación.

Este problema se aplica a varias situaciones entre las que se encuentran las siguientes:

i) 10 parejas, formadas cada una por un hombre y una mujer, llegan a una fiesta y en ella se forman 10 nuevas parejas hombre-mujer al azar. ¿Cuál es la probabilidad de que ninguna persona quede con la pareja que iba?

ii) Se tienen dos juegos de cartas con 52 *cartas cada uno; se van sacando simultáneamente una carta de cada juego hasta agotar las cartas. ¿Cuál es la probabilidad de que en ninguna ocasión salgan dos cartas iguales?*

iii) Una persona escribe n cartas dirigidas a n personas distintas, mete cada carta en un sobre y, finalmente, escribe al azar las direcciones de las n personas, una en cada sobre. ¿Cuál es la probabilidad de que por lo menos una de las cartas llegue a la persona para quien fue escrita?

EJEMPLO 3.8. *Dos personas P y Q juegan a lanzar consecutivamente una moneda balanceada con la condición de que cada vez que se obtenga cara, P gana un punto, mientras que Q lo gana cuando se obtiene cruz. Cada persona apuesta una cantidad x y convienen en que gana el juego quien obtenga primero n puntos. Supongamos que cuando al jugador P le faltan r puntos para ganar y al jugador Q s puntos, se les pierde la moneda y no pueden continuar el juego. ¿Cómo deben repartirse las apuestas en ese momento de manera que se tome en cuenta correctamente el número de puntos que lleva ganado cada uno?*

Solución

Se trata de calcular, en la situación dada, la probabilidad que cada jugador tiene de ganar y entonces la repartición de apuestas debe ser en la misma proporción que las respectivas probabilidades.

Para calcular la probabilidad buscada, se pueden obtener todas las posibles combinaciones de juegos ganados y perdidos que conducen a la victoria de cada jugador; la probabilidad de cada una de ellas puede obtenerse mediante la propiedad de independencia y entonces la probabilidad que cada jugador tiene de ganar se obtiene aplicando la propiedad de la aditividad finita. Sin embargo, si bien la lógica de este método es bastante simple, no conduce fácilmente a una fórmula que resuelva el problema de manera general. Vamos a ver que planteando el problema como un problema de muestreo se obtiene una solución general y muy simple.

Observemos que en la situación dada, para que alguno de los dos gane el juego necesitarían lanzar la moneda a lo más $r + s - 1$ veces. Así que consideremos el experimento consistente en jugar ese número de lanzamientos. Este experimento tiene un total de 2^{r+s-1} posibles resultados, los cuales son equiprobables y cada uno de ellos se puede representar mediante una serie de letras del tipo $(P, P, Q, B, Q, Q, \ldots)$, en donde P y Q significa que el jugador P y Q ganan el correspondiente lanzamiento, respectivamente. De estas series, aquellas que tengan por lo menos r veces la letra P representan resultados favorables al triunfo de P y aquellas que tengan por lo menos s veces la letra Q representan resultados favorables al triunfo de Q.

Pero de entre todas las posibles series de $r+s-1$ letras, el número de aquéllas en las que hay exactamente k veces la letra P es igual al número de combinaciones de $r+s-1$ objetos tomadas de k en k, es decir $\binom{r+s-1}{k}$. Por lo tanto, el total de resultados favorables al triunfo de P es:

$$\binom{r+s-1}{r} + \binom{r+s-1}{r+1} + \cdots + \binom{r+s-1}{r+s-1}$$

y el total de resultados favorables al triunfo de Q es:

$$\binom{r+s-1}{s} + \binom{r+s-1}{s+1} + \cdots + \binom{r+s-1}{r+s-1}$$

Entonces si p es la probabilidad de que P gane y q la probabilidad de que Q gane, tenemos:

$$p = \frac{\binom{r+s-1}{r} + \binom{r+s-1}{r+1} + \cdots + \binom{r+s-1}{r+s-1}}{2^{r+s-1}}$$

$$q = \frac{\binom{r+s-1}{s} + \binom{r+s-1}{s+1} + \cdots + \binom{r+s-1}{r+s-1}}{2^{r+s-1}}$$

Así que, la repartición de las apuestas debe ser como $p : q$.

Este problema tiene una gran importancia desde el punto de vista histórico pues con base en él se desarrollaron algunos métodos de solución de problemas de probabilidad. El método que dimos aquí es debido a Fermat (1654), inmediatamente después, Pascal, usando otro método, encontró la siguiente regla, la cual es sorprendente por su simplicidad y elegancia.

La regla de Pascal está basada en el llamado triángulo aritmético, el cual se construye de la siguiente manera:

$$\begin{array}{ccccccccccccccccc}
 & & & & & & & & 1 & & & & & & & & \\
 & & & & & & & 1 & & 1 & & & & & & & \\
 & & & & & & 1 & & 2 & & 1 & & & & & & \\
 & & & & & 1 & & 3 & & 3 & & 1 & & & & & \\
 & & & & 1 & & 4 & & 6 & & 4 & & 1 & & & & \\
 & & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & & \\
 & & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & & \\
 & 1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1 & \\
1 & & 8 & & 28 & & 56 & & 70 & & 56 & & 28 & & 8 & & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

Cada elemento de un renglón se obtiene sumando los dos elementos adyacentes del renglón anterior.

Entonces si a P le faltan r puntos y a Q s puntos, para determinar como debe ser la repartición de las apuestas, tómese del triángulo aritmético el renglón que tenga $r+s$ elementos. Si H es la suma de los primeros s elementos y G es la suma de los r últimos, la repartición de las apuestas entre P y G debe ser en la proporción $H : G$.

Para demostrar lo anterior observemos que $\binom{1}{0} = \binom{1}{1} = 1$, así que, gracias a la relación $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$, el triángulo aritmético puede expresarse de la siguiente manera:

$$
\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & \binom{1}{0} & & \binom{1}{1} & & & & & \\
 & & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & & \\
 & & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & & \\
 & & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & & \\
 & \binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5} & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}
$$

De esta manera, si consideramos el renglón del triángulo aritmético con $r+s$ elementos, la suma de los primeros s elementos resulta ser $\binom{r+s-1}{0} + \binom{r+s-1}{1} + \cdots + \binom{r+s-1}{s-1} = \binom{r+s-1}{r} + \binom{r+s-1}{r+1} + \cdots + \binom{r+s-1}{r+s-1}$ y la suma de los últimos r elementos resulta ser $\binom{r+s-1}{s} + \binom{r+s-1}{s+1} + \cdots + \binom{r+s-1}{r+s-1}$, y entonces, efectivamente, de acuerdo al resultado obtenido previamente, la repartición de las apuestas debe realizarse proporcionalmente a estas sumas.

Ejemplo 3.9 (**Probabilidades de tipo binomial**). *Una población contiene r objetos de tipo I y s objetos de tipo II. Se toma una muestra aleatoria con reemplazo de tamaño n y queremos calcular la probabilidad de que la muestra contenga exactamente k objetos de tipo I, en donde $k \leq n$.*

Podemos razonar como sigue: k objetos de tipo I y $n-k$ de tipo II pueden obtenerse de diferentes maneras, por ejemplo, si los primeros k son de tipo I y los restantes de tipo II, si los primeros $k-1$ y el $k+1$ son de tipo I y los restantes de tipo II, etc., es decir, hay tantas maneras como maneras hay de elegir k lugares de entre los n, lo cual es igual a $\binom{n}{k}$. De acuerdo con la regla del producto, cada uno de estos casos tiene una probabilidad igual a $\left(\frac{r}{r+s}\right)^k \left(\frac{s}{r+s}\right)^{n-k}$, de manera que si llamamos A_k al evento en consideración, se obtiene finalmente, para $k \in \{0, 1, 2, \ldots, n\}$:

$$P(A_k) = \binom{n}{k} \left(\frac{r}{r+s}\right)^k \left(\frac{s}{r+s}\right)^{n-k}$$

Ejemplo 3.10 (**Probabilidades de tipo hipergeométrico**). *Una población contiene r objetos de tipo I y s objetos de tipo II. Se toma una muestra aleatoria sin* *reemplazo de tamaño n y queremos calcular la probabilidad de que la muestra contenga exactamente k objetos de tipo I, en donde $k \leq n$, $k \leq r$ y $n - k \leq s$.*

Como ya se mencionó antes, se obtiene el mismo resultado independientemente de que el muestreo sea ordenado o no. Asumiendo que el muestreo es no ordenado, tenemos un total de $\binom{r+s}{n}$ posibles resultados, los cuales son equiprobables. De éstos, el total de casos en los cuales se obtienen k objetos de tipo I y $n-k$ objetos de tipo II es igual a $\binom{r}{k}\binom{s}{n-k}$. De manera que, llamando B_k al evento en consideración, se obtiene:

$$P(B_k) = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}$$

▲

Es interesante comparar las probabilidades obtenidas en los dos casos anteriores. Esperaríamos que si r y s son grandes, estas probabilidades se parezcan. En efecto, se puede demostrar la siguiente relación:

$$\left(1 - \frac{k}{r}\right)^k \left(1 - \frac{n-k}{s}\right)^{n-k} P(A_k) < P(B_k) < \left(1 - \frac{n}{r+s}\right)^{-n} P(A_k)$$

Así que si r y s son grandes comparados con n y k, $P(A_k) \approx P(B_k)$. Es decir, nuevamente obtenemos que, para poblaciones grandes, prácticamente no hay diferencia entre el muestreo con reemplazo y el muestreo sin reemplazo.

3.4. Coeficientes binomiales

La cantidad $\binom{N}{n}$ aparece en numerosos lugares, de ahí la importancia de estudiar algunas de sus propiedades.

En primer lugar, la definición se puede extender como sigue:

Definición 3.11 (**Coeficientes binomiales**). *Sean $z \in \mathbb{R}$ y $m \in \mathbb{Z}$, entonces definimos:*

$$\binom{z}{m} = \begin{cases} \frac{z(z-1)(z-2)\cdots(z-m+1)}{m!} & \text{si } m > 0 \\ 1 & \text{si } m = 0 \\ 0 & \text{si } m < 0 \end{cases}$$

También definimos $0! = 1$.

Obviamente, cuando z y m son enteros no negativos, entonces:

$$\binom{z}{m} = \begin{cases} \frac{z!}{m!(z-m)!} & \text{si } m \leq z \\ 0 & \text{si } z < m \end{cases}$$

Además, en este caso se tiene $\binom{z}{m} = \binom{z}{z-m}$.

PROPOSICIÓN 3.12. *Para cualesquiera $z \in \mathbb{R}$ y $m \in \mathbb{Z}$, se tiene:*

a) $\binom{z}{m} = (-1)^m \binom{m-1-z}{m}$

b) $\binom{z}{m} + \binom{z}{m+1} = \binom{z+1}{m+1}$

Demostración

a. $\binom{z}{m} = \frac{z(z-1)(z-2)\cdots(z-m+1)}{m!} = (-1)^m \frac{-z(1-z)(2-z)\cdots(m-1-z)}{m!} = (-1)^m \binom{m-1-z}{m}$

b. Para $m > 0$, se tiene:

$\binom{z}{m} + \binom{z}{m+1} = \frac{z(z-1)(z-2)\cdots(z-m+1)}{m!} + \frac{z(z-1)(z-2)\cdots(z-m)}{(m+1)!}$

$= \frac{z(z-1)(z-2)\cdots(z-m+1)}{m!} \left[1 + \frac{z-m}{m+1}\right] = \frac{(z+1)z(z-1)(z-2)\cdots(z-m+1)}{(m+1)!} = \binom{z+1}{m+1}$

Para $m = 0$, $\binom{z}{m} + \binom{z}{m+1} = z + 1 = \binom{z+1}{m+1}$

Para $m = -1$, $\binom{z}{m} + \binom{z}{m+1} = 1 = \binom{z+1}{m+1}$

Para $m < -1$, $\binom{z}{m} + \binom{z}{m+1} = 0 = \binom{z+1}{m+1}$

■

PROPOSICIÓN 3.13 (**Teorema del binomio**). *Sean $t, x \in \mathbb{R}$, entonces:*

$$(1+t)^z = \begin{cases} \sum_{k=0}^{z} \binom{z}{k} t^k \text{ para cualquier } t \in \mathbb{R} & \text{si } z \in \{0, 1, \ldots\} \\ \sum_{k=0}^{\infty} \binom{z}{k} t^k \text{ para } |t| < 1 & \text{en otro caso} \end{cases}$$

Demostración

Desarrollando la función $f(z) = (1+t)^z$ en serie de Taylor alrededor de 0, se obtiene:

$(1+t)^z = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} t^k = 1 + zt + \frac{z(z-1)}{2!} t^2 + \frac{z(z-1)(z-2)}{3!} t^3 + \cdots$

$= \binom{z}{0} + \binom{z}{1} t + \binom{z}{2} t^2 + \binom{z}{3} t^3 + \cdots = \sum_{k=0}^{\infty} \binom{z}{k} t^k$

Si $z \in \{0, 1, \ldots\}$, entonces $\binom{z}{k} = 0$ para cualquier entero $k > z$, así que: la serie converge para cualquier $t \in \mathbb{R}$ y $\sum_{k=0}^{\infty} \binom{z}{k} t^k = \sum_{k=0}^{z} \binom{z}{k} t^k$.

Si $z \notin \{0, 1, \ldots\}$, entonces el radio de convergencia R de la serie está dado por:

$R = \text{lím}_{k \rightsquigarrow \infty} \left| \frac{\binom{z}{k}}{\binom{z}{k+1}} \right| = \text{lím}_{k \rightsquigarrow \infty} \left| \frac{\frac{z(z-1)\cdots(z-k+1)}{k!}}{\frac{z(z-1)\cdots(z-k)}{(k+1)!}} \right| = \text{lím}_{k \rightsquigarrow \infty} \left| \frac{k+1}{z-k} \right| = 1$

Así que la serie converge para $|t| < 1$.

■

Corolario 3.14 (**Teorema del binomio de Newton**). *Si a y b son dos números reales cualesquiera y $n \in \mathbb{N}$, entonces:*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Corolario 3.15. *Si a y b son dos números reales tales que $\left|\frac{a}{b}\right| < 1$ y $z \in \mathbb{R}$, entonces:*

$$(a+b)^z = \sum_{k=0}^{\infty} \binom{z}{k} a^k b^{z-k}$$

Proposición 3.16. *Sean $n, m \in \mathbb{N}$, entonces:*

$\sum_{k=1}^{n} k^m \binom{n}{k} = n2^{n-1} + n \sum_{j=1}^{m-1} \binom{m-1}{j} \sum_{k=1}^{n-1} k^j \binom{n-1}{k}$

Demostración

$\sum_{k=1}^{n} k^m \binom{n}{k} = \sum_{k=1}^{n} k^{m-1} \frac{n!}{(k-1)!(n-k)!} = n \sum_{k=1}^{n} k^{m-1} \binom{n-1}{k-1}$

$= n \sum_{k=1}^{n} [(k-1)+1]^{m-1} \binom{n-1}{k-1} = n \sum_{k=1}^{n} \left[1 + \sum_{j=1}^{m-1} \binom{m-1}{j} (k-1)^j\right] \binom{n-1}{k-1}$

$= n \sum_{k=1}^{n} \binom{n-1}{k-1} + n \sum_{k=1}^{n} \sum_{j=1}^{m-1} \binom{m-1}{j} (k-1)^j \binom{n-1}{k-1}$

$= n \sum_{k=0}^{n-1} \binom{n-1}{k} + n \sum_{k=1}^{n-1} \sum_{j=1}^{m-1} \binom{m-1}{j} k^j \binom{n-1}{k}$

$= n2^{n-1} + n \sum_{j=1}^{m-1} \binom{m-1}{j} \sum_{k=1}^{n-1} k^j \binom{n-1}{k}$

■

Corolario 3.17. *Sea $n \in \mathbb{N}$, entonces:*

a) $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

b) $\sum_{k=1}^{n} k \binom{n}{k} = n2^{n-1}$

c) $\sum_{k=1}^{n} k^2 \binom{n}{k} = (n^2+n)2^{n-2}$

d) $\sum_{k=1}^{n} k^3 \binom{n}{k} = (n^3+3n^2)\, 2^{n-3}$

Demostración

a. Resulta de aplicar el teorema del binomio de Newton para $a = b = 1$.

b. $\sum_{k=1}^{n} k \binom{n}{k} = n2^{n-1} + n \sum_{j=1}^{0} \binom{m-1}{j} \sum_{k=1}^{n-1} k^j \binom{n-1}{k} = n2^{n-1}$

c. $\sum_{k=1}^{n} k^2\binom{n}{k} = n2^{n-1} + n\sum_{k=1}^{n-1} k\binom{n-1}{k} = n2^{n-1} + n(n-1)2^{n-2} = (n^2+n)2^{n-2}$

d. $\sum_{k=1}^{n} k^3\binom{n}{k} = n2^{n-1} + n\left[2\sum_{k=1}^{n-1} k\binom{n-1}{k} + \sum_{k=1}^{n-1} k^2\binom{n-1}{k}\right]$

$= n2^{n-1} + 2n(n-1)2^{n-2} + n^2(n-1)2^{n-3} = (n^3+3n^2)\,2^{n-3}$

■

Proposición 3.18. $\sum_{k=1}^{n}(-1)^k k^m\binom{n}{k} = 0$, *para cualesquiera* $n, m \in \mathbb{N}$, *con* $n > m$.

Demostración

La demostración es por inducción sobre m.

Para $n > 1$, se tiene:

$\sum_{k=1}^{n}(-1)^k k\binom{n}{k} = \sum_{k=1}^{n}(-1)^k \frac{n!}{(k-1)!(n-k)!} = -n\sum_{k=1}^{n}(-1)^{k-1}\binom{n-1}{k-1}$

$= -n\sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k} = -n(-1+1)^{n-1} = 0$

Supongamos ahora que $\sum_{k=0}^{n}(-1)^k k^j\binom{n}{k} = 0$ para cualquier $j \le m$ y cualquier $n > j$, entonces, para $n > m+1$, se tiene:

$\sum_{k=0}^{n}(-1)^k k^{m+1}\binom{n}{k} = \sum_{k=1}^{n}(-1)^k\left[(k-1)+1\right]^m \frac{n!}{(k-1)!(n-k)!}$

$= -n\sum_{k=1}^{n}(-1)^{k-1}\sum_{j=0}^{m}\binom{m}{j}(k-1)^j\binom{n-1}{k-1} = -n\sum_{j=0}^{m}\binom{m}{j}\sum_{k=0}^{n-1}(-1)^k k^j\binom{n-1}{k} = 0$

■

EJERCICIOS

Ejercicio 3.1. *Demuestre las siguientes relaciones:*

a) $\sum_{k=n}^{m}\binom{k}{n} = \binom{m+1}{n+1}$

b) $\sum_{k=0}^{n}\binom{k+r-1}{k} = \binom{n+r}{n}$

c) $\sum_{k=0}^{r}\binom{n}{k}\binom{m}{r-k} = \binom{n+m}{r}$

d) $\sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n}$

e) $\sum_{k=m}^{n}\binom{n}{k}\binom{k}{m} = \binom{n}{m}2^{n-m}$

f) $\sum_{k=0}^{m}\binom{n}{k}\binom{n-k}{m-k} = \binom{n}{m}2^m$

g) $\sum_{k=m}^{n}(-1)^k\binom{n}{k}\binom{k}{m}=0$

h) $\sum_{k=0}^{m}(-1)^k\binom{n}{k}\binom{n-k}{m-k}=0$

en donde $m,n,r\in\mathbb{N}$.

Ejercicio 3.2. *Demuestre que:*

$\sum_{k=1}^{n}k^m=\frac{1}{m+1}\left[(n+1)^{m+1}-(n+1)-\sum_{j=1}^{m-1}\binom{m+1}{j}\sum_{k=1}^{n}k^j\right]$

para cualesquiera $n,m\in\mathbb{N}$.

***Sugerencia**: Utilice la relación* $k^{m+1}=[(k-1)+1]^{m+1}=1+\sum_{j=1}^{m+1}\binom{m+1}{j}(k-1)^j$

Ejercicio 3.3. *Utilice el resultado del ejercicio anterior para demostrar las siguientes relaciones:*

a) $\sum_{k=1}^{n}k=\frac{1}{2}n(n+1)$

b) $\sum_{k=1}^{n}k^2=\frac{1}{6}n\,(n+1)\,(2n+1)$

c) $\sum_{k=1}^{n}k^3=\frac{1}{4}n^2\,(n+1)^2$

d) $\sum_{k=1}^{n}k^4=\frac{1}{30}n\,(n+1)\,(2n+1)\,(3n^2+3n-1)$

en donde $n\in\mathbb{N}$.

Ejercicio 3.4. *De una población de* N *objetos,* $b_1,\ldots,b_N$*, se toma una muestra aleatoria ordenada sin reemplazo de tamaño* n*. Dado* $k\in\{1,\ldots,N\}$*, calcule la probabilidad de que* b_k *se encuentre en la muestra.*

Ejercicio 3.5. *Se tienen* 2 *urnas, cada una de las cuales contiene* 10 *bolas numeradas del* 1 *al* 10*. Un experimento aleatorio consiste en seleccionar al azar una bola de cada urna. Calcule la probabilidad de que los números de las dos bolas seleccionadas difieran por* 2 *o más.*

Ejercicio 3.6. *Un experimento aleatorio consiste en lanzar* 3 *dados sobre una mesa, uno de los dados es rojo, otro azul y el otro blanco. ¿Cuál es la probabilidad de que a) los* 3 *números que se obtienen sean distintos? b) b) el número que se obtiene con el dado azul sea igual a la suma de los números que se obtienen con los otros dos dados?*

Ejercicio 3.7. *Una urna contiene* 8 *bolas rojas,* 7 *bolas negras y* 10 *bolas blancas. Un experimento aleatorio consiste en seleccionar, al azar y con reemplazo,* 3 *bolas de la urna. Calcule la probabilidad de que a) las* 3 *bolas seleccionadas sean del mismo color y b) por lo menos una de las bolas seleccionadas sea blanca.*

Ejercicio 3.8. 5 *personas suben a un elevador en la planta baja de un edificio de* 10 *pisos. Supongamos que cada persona puede bajar en cualquiera de los* 10 *pisos con la misma probabilidad. Calcule la probabilidad de que a) por lo menos una persona baje del elevador después del piso* 5 *y b) las* 5 *personas bajen del elevador en pisos distintos.*

Ejercicio 3.9. *Sabiendo que al lanzar un par de dados se obtiene una suma igual a* 6*, ¿cuál es la probabilidad de que uno de los números que se obtienen sea el* 1*?*

Ejercicio 3.10. *Un experimento aleatorio consiste en lanzar un par de dados. Sabiendo que como resultado del experimento se obtienen dos números distintos, ¿cuál es la probabilidad de que uno de esos números sea el* 4*?*

Ejercicio 3.11. *Un experimento aleatorio consiste en seleccionar, al azar y con reemplazo,* 2 *bolas de una urna, la cual contiene* 3 *bolas blancas y* 2 *bolas negras. Si se sabe que las dos bolas seleccionadas son del mismo color, ¿cuál es la probabilidad de que las dos sean blancas?*

Ejercicio 3.12. *Un experimento aleatorio consiste en lanzar un dado* 10 *veces. Si se sabe que se obtiene por lo menos un* 6*, ¿cuál es la probabilidad de obtener dos o más* 6*'s?*

Ejercicio 3.13. *Un dado se lanza* 10 *veces. Suponiendo que en* 6 *ocasiones se obtiene un número entre* 1 *y* 4 *(inclusive), ¿cuál es la probabilidad de que se obtenga por lo menos una vez el número* 1*?*

Ejercicio 3.14. *Un dado es lanzado* 3 *veces. Si se sabe que se obtiene al menos una vez el número* 6*, ¿cuál es la probabilidad de que se obtenga exactamente una vez el número* 6*?*

Ejercicio 3.15. *Un dado se lanza* 5 *veces. Sabiendo que los* 5 *números que se obtienen son distintos, ¿cuál es la probabilidad de que entre ellos no esté el número* 6*?*

Ejercicio 3.16. *Se van colocando al azar bolas, una a una, en cualquiera de* N *cajas, hasta que alguna caja llegue a tener* 2 *bolas.¿Cuál es la probabilidad de que el proceso termine en el paso* n*?*

Ejercicio 3.17. *Se seleccionan, al azar y con reemplazo,* 3 *tarjetas de una urna que contiene* N *tarjetas numeradas del* 1 *al* N*. Encuentre la probabilidad de que el menor de los números seleccionados sea igual a* 2*.*

Ejercicio 3.18. *Se seleccionan, al azar y con reemplazo,* 4 *tarjetas de una urna que contiene* N *tarjetas numeradas del* 1 *al* N*. Encuentre la probabilidad de que el mayor de los números seleccionados sea igual a* N*.*

Ejercicio 3.19. *Se seleccionan, al azar y con reemplazo,* 10 *tarjetas de una urna que contiene* N *tarjetas numeradas del* 1 *al* N*. Encuentre la probabilidad de que a) el menor de los números seleccionados sea igual a* k *y b) el mayor de los números seleccionados sea igual a* k*, en donde* $k \leq N$*.*

Ejercicio 3.20. *Antonio y Luis forman parte de un grupo de* 12 *personas, las cuales ocupan al azar* 12 *sillas que se encuentren formando un círculo. ¿Cuál es la probabilidad de que queden exactamente* 4 *personas entre Antonio y Luis en el arco que va de Antonio a Luis en dirección positiva?*

Ejercicio 3.21. *Antonio y Luis forman parte de un grupo de* n *personas* $(n > 2)$*, las cuales ocupan al azar* n *sillas que se encuentren formando un círculo. ¿Cuál es la probabilidad de que a) Antonio y Luis queden sentados uno al lado del otro? y b) queden exactamente* k *personas entre Antonio y Luis?*

Ejercicio 3.22. n *personas se sientan al azar en* n *sillas alrededor de una mesa redonda. Encuentre la probabilidad de que las personas A, B y C queden juntas, con A a la derecha de B y C a la izquierda de B.*

Ejercicio 3.23. *Los números* $1, \ldots, n$ $(n \geq 3)$ *se colocan uno después del otro en un orden aleatorio. Encuentre la probabilidad de que los números* 1*,* 2 *y* 3 *queden colocados juntos y ordenados en forma creciente.*

Ejercicio 3.24. *Juan y Pedro forman parte de un grupo de* 10 *personas, las cuales ocupan al azar* 10 *sillas que se encuentren en hilera. ¿Cuál es la probabilidad de que queden exactamente* 2 *personas entre Juan y Pedro?*

Ejercicio 3.25. 5 *hombres y* 5 *mujeres se sientan al azar en* 10 *sillas colocadas en hilera. Encuentre la probabilidad de que por lo menos una de las personas quede sentada junto a otra del mismo sexo.*

Ejercicio 3.26. *Un tablero circular se divide en tres zonas acotadas por círculos concéntricos de radios* $\frac{1}{3}$*,* $\frac{1}{2}$ *y* 1*, respectivamente. Si se hacen tres disparos al azar sobre el tablero, ¿cuál es la probabilidad de que cada zona reciba uno de los disparos?*

Ejercicio 3.27. *Cada una de* n *bolas se coloca al azar en una de* n *cajas. Encuentre la probabilidad de que exactamente una de las cajas quede vacía.*

Ejercicio 3.28. n *bolas se encuentren colocadas en* n *cajas, una en cada caja. Un experimento consiste en sacar las bolas de la caja y en colocarlas después, al azar, nuevamente una en cada caja. Calcule la probabilidad de que al realizar el experimento, exactamente* k *bolas queden en las cajas en las que estaban originalmente.*

Ejercicio 3.29. 10 *parejas, formadas cada una por un hombre y una mujer, llegan a una fiesta y en ella se forman* 10 *nuevas parejas hombre-mujer al azar. ¿Cuál es la probabilidad de que a) ninguna persona quede con la pareja que iba? y b) a lo más* 2 *parejas queden formadas como lo estaban originalmente?*

EJERCICIO 3.30. *Una persona escribe n cartas dirigidas a n personas distintas, mete cada carta en un sobre y, finalmente, escribe al azar las direcciones de las n personas, una en cada sobre. Encuentre la probabilidad de que por lo menos una de las cartas llegue a la persona para quien fue escrita.*

EJERCICIO 3.31. *Una urna contiene 4 bolas blancas, 4 rojas y 4 azules. Se van seleccionando bolas de la urna, una a una y sin reemplazo, hasta obtener en forma consecutiva dos bolas del mismo color, o hasta que se agoten las bolas de la urna. Encuentre la probabilidad de que este proceso termine en la quinta elección.*

EJERCICIO 3.32. *Tres urnas contienen bolas blancas y negras; la primera 2 blancas y 3 negras, la segunda 2 blancas y 2 negras y la tercera 3 blancas y 1 negra. Se transfiere una bola de la primera a la segunda urna, después una de la segunda a la tercera y, finalmente, una de la tercera a la primera, siendo todas las transferencias al azar. ¿Qué composición de bolas en la primera urna, después de las transferencias, es la más probable?*

EJERCICIO 3.33. *Una urna contiene r bolas rojas y b bolas blancas. Se elige al azar una bola de la urna, después, sin reemplazar la primera, se elige otra y así sucesivamente hasta sacar n bolas. Calcule la probabilidad de que en el j-ésimo paso se saque una bola roja sabiendo que la muestra contiene s bolas rojas.*

EJERCICIO 3.34. *En una urna hay N tarjetas numeradas del 1 al N. Se van seleccionando al azar las tarjetas de la urna, una a una y sin reemplazo, hasta que se seleccionan todas. Para $k \in \{1, \ldots, N\}$, sea A_k el evento 'el número de la k-ésima tarjeta seleccionada es mayor que los números de las seleccionadas previamente'. Encuentre la probabilidad de cada evento A_k y demuestre que los eventos $A_1, \ldots, A_n$ son independientes.*

EJERCICIO 3.35. *Una urna contiene 20 bolas numeradas del 1 al 20. Se seleccionan 2 bolas, al azar y sin reemplazo. Encuentre la probabilidad de que los dos números de las bolas seleccionadas difieran por 3 o más.*

EJERCICIO 3.36. *Se eligen al azar dos números del conjunto $\{1, \ldots, n\}$. Encuentre la probabilidad de que uno de los números seleccionados sea menor que k y el otro sea mayor que k, en donde $1 < k < n$.*

EJERCICIO 3.37. *Un experimento aleatorio consiste en seleccionar, al azar y sin reemplazo, 3 bolas de una urna, la cual contiene 6 bolas blancas y 4 bolas negras. Si se sabe que las tres bolas seleccionadas son del mismo color, ¿cuál es la probabilidad de que las tres sean blancas?*

EJERCICIO 3.38. *De una urna que contiene 6 bolas blancas y 4 bolas rojas se seleccionan, sin reemplazo, 4 bolas al azar y se transfieren a una segunda urna, la cual se encuentre vacía. Inmediatamente después, se seleccionan, al azar y sin reemplazo, dos bolas de la segunda urna. Sabiendo que entre las 4 bolas que se transfieren hay* *por lo menos* 2 *blancas, calcule la probabilidad de que las dos bolas seleccionadas de la segunda urna sean ambas blancas.*

EJERCICIO 3.39. *Un experimento aleatorio consiste en seleccionar, al azar y sin reemplazo,* 3 *bolas de una urna, la cual contiene* 6 *bolas negras y* 4 *bolas blancas. Si se sabe que al menos una de las bolas seleccionadas es negra, ¿cuál es la probabilidad de que las* 3 *sean negras?*

EJERCICIO 3.40. *De una urna que contiene* 4 *bolas rojas y* 6 *blancas se seleccionan, al azar y sin reemplazo, dos bolas; inmediatamente después se selecciona al azar una de las bolas restantes. ¿Cuál es la probabilidad de que la tercera bola seleccionada sea roja si se sabe que entre las dos primeras hay por lo menos una roja?*

EJERCICIO 3.41. *Cada una de dos urnas contiene* N *bolas numeradas del* 1 *al* N*. De cada una de las cajas se toma una muestra de tamaño* n *sin reemplazo. Calcule la probabilidad de obtener exactamente* k *bolas con el mismo número.*

EJERCICIO 3.42. *Cada una de* N *personas estaciona su automóvil en uno de* N *lugares consecutivos para ser atendidos en una oficina. El automóvil de una persona* A *no queda en ninguno de los extremos y antes de* A *son atendidas* n *personas, cada una de las cuales retira su auto. Asumiendo que las* N *personas son atendidas al azar, ¿cuál es la probabilidad de que a)* A *encuentre desocupados los dos lugares contiguos a su auto? y b)* A *encuentre ocupados los dos lugares contiguos a su auto?*

EJERCICIO 3.43. *Pedro es un arquero y se sabe que, al disparar una flecha, la probabilidad de que ésta de en el blanco es igual a* $\frac{9}{10}$*. Si Pedro realiza* 6 *lanzamientos de manera independiente, a) ¿cuál es la probabilidad de que acierte en el blanco por lo menos una vez?. b) Sabiendo que, de los* 6 *disparos, acierta en el blanco por lo menos una vez, ¿cuál es la probabilidad de que acierte en el blanco por lo menos dos veces?*

EJERCICIO 3.44. *Una urna contiene* 6 *bolas rojas y* 4 *bolas blancas. Un experimento aleatorio consiste en seleccionar, al azar y con reemplazo,* 20 *bolas de la urna. Encuentre la probabilidad de que se obtengan por lo menos* 4 *bolas rojas en la muestra.*

EJERCICIO 3.45. *Una urna contiene* 8 *tarjetas marcadas con números positivos y* 10 *tarjetas marcadas con números negativos. Un experimento aleatorio consiste en seleccionar, al azar y con reemplazo,* 4 *tarjetas de la urna. Encuentre la probabilidad de que el producto de los* 4 *números que se obtienen sea positivo.*

EJERCICIO 3.46. *Calcule la probabilidad de obtener exactamente* k *veces* 6 *en* n *lanzamientos de un dado*[2]*.*

[2]Este problema fue uno de los primeros problemas de probabilidad que se plantearon. En las primeras soluciones no era evidente la relación que había con el desarrollo de un binomio, ésta fue encontrada más tarde. Gracias a esta relación, este tipo de probabilidades adquirió una gran importancia en el desarrollo posterior de la Teoría de la Probabilidad (como puede verse en el capítulo siguiente) y actualmente sigue siendo uno de los tipos principales.

EJERCICIO 3.47. *Un lote contiene n artículos. Si se sabe que r artículos son defectuosos y se inspeccionan uno por uno en un orden aleatorio, ¿cuál es la probabilidad de que el k-ésimo artículo ($k \geq r$) inspeccionado sea el último defectuoso en el lote?*

EJERCICIO 3.48. *Una urna contiene 8 bolas rojas, 7 bolas negras y 10 bolas blancas. Un experimento aleatorio consiste en seleccionar, al azar y sin reemplazo, 3 bolas de la urna. Calcule la probabilidad de que a) las 3 bolas seleccionadas sean del mismo color y b) por lo menos una de las bolas seleccionadas sea blanca.*

EJERCICIO 3.49. *De una urna que contiene 10 bolas rojas, 10 bolas negras y 10 bolas blancas, se seleccionan 2 bolas al azar y sin reemplazo. Calcule la probabilidad de que las dos bolas seleccionadas sean de distinto color.*

EJERCICIO 3.50. *Una urna contiene 6 tarjetas marcadas con números positivos y 8 tarjetas marcadas con números negativos. Un experimento aleatorio consiste en seleccionar, al azar y sin reemplazo, 4 tarjetas de la urna. Encuentre la probabilidad de que el producto de los 4 números que se obtienen sea positivo.*

EJERCICIO 3.51. *Una urna contiene 16 bolas rojas y 14 bolas blancas. Un experimento aleatorio consiste en seleccionar, al azar y sin reemplazo, 10 bolas de la urna. Encuentre la probabilidad de que se obtengan a lo más 4 bolas rojas.*

EJERCICIO 3.52. *En una fábrica hay 35 trabajadores por contrato, de los cuales 20 laboran en el turno matutino, y 40 con definitividad, de los cuales 25 laboran en el turno matutino. En una asamblea de todos los trabajadores, se forma una comisión eligiendo 6 personas al azar. ¿Cuál es la probabilidad de que queden representados en la comisión los 4 sectores de trabajadores?*

EJERCICIO 3.53. *En una rifa se reparten n boletos, de los cuales m obtendrán algún premio. Encuentre la probabilidad de que una persona, que tiene k boletos, obtenga por lo menos un premio.*

EJERCICIO 3.54. *Un grupo de $2N$ niños y $2N$ niñas se divide en dos grupos de $2N$ personas cada uno. Encuentre la probabilidad de que cada grupo contenga el mismo número de niños que de niñas.*

EJERCICIO 3.55. *Una urna contiene 4 bolas rojas, 6 bolas negras y 8 bolas blancas. Un experimento aleatorio consiste en seleccionar, al azar y sin reemplazo, 4 bolas de la urna. Calcule la probabilidad de que se obtengan por lo menos 2 bolas blancas en la muestra.*

CAPÍTULO 4

COMBINANDO LAS REGLAS BÁSICAS

Sea que sobrevengan "espontáneamente" o que sean "inducidas artifialmente", las mutaciones aparecen siempre al azar. Nunca se encuentra alguna relación entre su producción y las condiciones externas, ninguna dirección dada por el medio.

Francois Jacob

Los eventos elementales iniciales que abren la via de la evolución a esos sistemas intensamente conservadores que son los seres vivos, son microscópicos y fortuitos... Pero una vez inscritos en la estructura del ADN, el accidente singular y como tal esencialmente impredecible será mecanicamente y fielmente replicado y traducido... Sacado del reino del azar puro, entra en el de la necesidad, el de las certidumbres más implacables.

Jacques Monod

4.1. Regla de la probabilidad total

El principal método para calcular probabilidades consiste en la utilización de las reglas de la suma y del producto. Si uno quiere calcular la probabilidad de un evento complejo, se busca expresar ese evento como una unión o una intersección de eventos más simples; de preferencia, eventos mutuamente excluyentes para el caso de una unión y eventos independientes para el caso de una intersección.

En la mayor parte de los casos, las reglas de la suma y del producto se utilizan simultáneamente, expresando el evento cuya probabilidad se quiere calcular, primero como una unión de intersecciones mutuamente excluyentes y, después, calculando la probabilidad de cada intersección utilizando la regla del producto.

Este método para calcular probabilidades se expresa en lo que se llama la regla de la probabilidad total, tema que se expone a continuación.

DEFINICIÓN 4.1 (**Partición de un espacio muestral**). *Se dice que una colección de eventos de probabilidad positiva* $A_1, \ldots, A_n$ *forman una partición del espacio muestral si éstos son mutuamente excluyentes y* $A_1 \cup \cdots \cup A_n = \Omega$.

PROPOSICIÓN 4.2 (**Regla de la probabilidad total**). *Sean* B *un evento cualquiera y* $A_1, \ldots, A_n$ *una partición del espacio muestral , entonces:*

$$P(B) = P(B \mid A_1)P(A_1) + \ldots + P(B \mid A_n)P(A_n)$$

Demostración

Los eventos $B \cap A_1, \ldots, B \cap A_n$ son mutuamente excluyentes y $B = (B \cap A_1) \cup \cdots \cup (B \cap A_n)$, así que, usando la regla de la suma primero y después la del producto, obtenemos:

$$P(B) = P\left[(B \cap A_1) \cup \cdots \cup (B \cap A_n)\right]$$

$$= P(B \cap A_1) + \cdots + P(B \cap A_n)$$

$$= P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_n)P(A_n)$$

■

EJEMPLO 4.3. *Una urna contiene* $2N$ *bolas numeradas del* 1 *al* $2N$. *Un experimento consiste en elegir al azar una bola de esa urna, dejándola fuera, y después, en elegir al azar una segunda. Calcule la probabilidad de que la suma de los números elegidos sea par.*

Solución

Consideremos los siguientes eventos:

B*: la suma de los dos números elegidos es par.*

A_1 *: el número de la primera elección es par.*

A_2 *: El número de la primera elección es impar.*

Usando la regla de la probabilidad total, obtenemos:

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) = \tfrac{N-1}{2N-1}\tfrac{N}{2N} + \tfrac{N-1}{2N-1}\tfrac{N}{2N} = \tfrac{N-1}{2N-1}$$

EJEMPLO 4.4. *Una urna contiene r bolas rojas y s bolas blancas; otra urna contiene a bolas rojas y b bolas blancas. Una bola elegida al azar es transferida de la primera urna a la segunda y, después de esto, se extrae la azar una bola de esta última. ¿Cuál es la probabilidad de que esta bola sea roja?*

Solución

Consideremos los siguientes eventos:

A_1 *: la bola transferida de la primera urna en la segunda es roja.*

A_2 *: la bola transferida de la primera urna en la segunda es blanca.*

B*: la bola extraída de la segunda urna es roja.*

Usando la regla de la probabilidad total, obtenemos:

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) = \frac{a+1}{a+b+1}\frac{r}{r+s} + \frac{a}{a+b+1}\frac{s}{r+s} = \frac{r(a+1)+as}{(r+s)(a+b+1)}$$

EJEMPLO 4.5. *Se tienen* 3 *escritorios, cada uno con dos cajones. El primer escritorio contiene una moneda de oro en cada cajón; el segundo contiene una moneda de oro en un cajón y una de plata en el otro; el tercero contiene una moneda de plata en cada cajón. Se elige un escritorio al azar, se abre un cajón y después el otro. Calcule la probabilidad de que el segundo cajón contenga una moneda de oro bajo la hipótesis de que el primero contiene una moneda de oro.*

Solución

Consideremos los eventos siguientes:

A_1 *: se elige el primer escritorio.*

A_2 *: se elige el segundo escritorio.*

A_3 *: se elige el tercer escritorio.*

B *: el primer cajón que se abre contiene una moneda de oro.*

A *: el segundo cajón contiene una moneda de oro.*

Entonces:

$$P(A \mid B) = \frac{P(A\cap B)}{P(B)} = \frac{P(A_1)}{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)} = \frac{\frac{1}{3}}{\frac{1}{3}+\frac{1}{2}\frac{1}{3}} = \frac{2}{3}$$

Podría pensarse que dado que el primer cajón que se abre contiene una moneda de oro, sólo hay dos posibilidades para el otro cajón, entonces la probabilidad de que *el segundo cajón contenga una moneda de oro debería ser de* $\frac{1}{2}$*. Sin embargo, ese razonamiento es incorrecto; en realidad hay* 3 *posibilidades (igualmente probables) de elegir un cajón con una moneda de oro y de ellas* 2 *son favorables al evento en consideración.*

Ejemplo 4.6 (**Urna de Polya**). *Una urna contiene bolas rojas y blancas. Un experimento aleatorio consiste en seleccionar sucesivamente una bola al azar de la urna de tal manera que, después de cada elección, la bola seleccionada se regresa a la urna junto con* c *bolas más del mismo color. Para* $n \in \mathbb{N}$*, sea* B_n *el evento 'la bola seleccionada en el paso* n *es blanca'. Vamos a demostrar que, cualquiera que sea la configuración de bolas en la urna y para cualquier* $n \in \mathbb{N}$*,* $P(B_n)$ *es igual a la proporción inicial de bolas blancas. La demostración es por inducción sobre* n*.*

Para $n = 1$*, el resultado es inmediato. Supongamos ahora que el resultado es cierto para* $n = k$ *y que el proceso se inicia con* r *bolas rojas y* b *bolas blancas, entonces,*

$P(B_{k+1}) = P(B_{k+1} \mid B_1)P(B_1) + P(B_{k+1} \mid B_1^c)P(B_1^c)$

$= \frac{b+c}{r+b+c}\frac{b}{r+b} + \frac{b}{r+b+c}\frac{r}{r+b} = \frac{b(b+c+r)}{(r+b+c)(r+b)} = \frac{b}{r+b}$

Ejemplo 4.7. *Una urna contiene* 2 *bolas rojas y* 3 *bolas blancas; una segunda contiene* 1 *bola roja y* 7 *bolas blancas y una tercera contiene* 3 *bolas rojas y* 4 *bolas blancas. Un experimento consiste en elegir al azar una de las urnas y, después de eso, en elegir al azar una bola de la urna elegida. Calcule la probabilidad de que a) esta última sea roja y b) la bola seleccionada sea roja sabiendo que se elige una de las dos primeras urnas.*

Solución

Consideremos los siguientes eventos:

A_1 *: la urna elegida es la primera.*

A_2 *: la urna elegida es la segunda.*

A_3 *: La urna elegida es la tercera.*

B *: la bola extraída es roja.*

a. Usando la regla de la probabilidad total, obtenemos:

$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + P(B \mid A_3)P(A_3)$

$= \frac{2}{5}\frac{1}{3} + \frac{1}{8}\frac{1}{3} + \frac{3}{7}\frac{1}{3} = \frac{89}{280}$

b. $P(B \mid A_1 \cup A_2) = \frac{P[B\cap(A_1\cup A_2)]}{P(A_1\cup A_2)} = \frac{P[(B\cap A_1)\cup(B\cap A_2)]}{P(A_1\cup A_2)} = \frac{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)}{P(A_1)+P(A_2)}$

$= P(B \mid A_1)\frac{P(A_1)}{P(A_1)+P(A_2)} + P(B \mid A_2)\frac{P(A_2)}{P(A_1)+P(A_2)} = \frac{2}{5}\frac{1}{2} + \frac{1}{8}\frac{1}{2} = \frac{21}{80}$

▲

Obsérvese que en el ejemplo anterior, $P(B \mid A_1 \cup A_2)$ tiene una interpretación natural, pues dada la ocurrencia de $A_1 \cup A_2$, la probabilidad del evento B se calcula olvidándonos de la tercera urna y suponiendo que el experimento consiste primero en elegir una de las dos primeras urnas y, después de eso, en extraer al azar una bola de la urna elegida. Este resultado puede generalizarse dándonos una interpretación de la probabilidad condicional en caso análogos al considerado en el ejemplo.

Sea $\mathcal{E}$ un experimento aleatorio, $A_1, \ldots, A_r$ r eventos mutuamente excluyentes relativos a $\mathcal{E}$ y B otro evento relativo a $\mathcal{E}$. Tenemos entonces:

$P(B \mid A_1 \cup \cdots \cup A_r) = \frac{P[B\cap(A_1\cup\cdots\cup A_r)]}{P(A_1\cup\cdots\cup A_r)} = \frac{P[(B\cap A_1)\cup\cdots\cup(B\cap A_r)]}{P(A_1\cup\cdots\cup A_r)}$

$= \frac{P(B|A_1)P(A_1)+\cdots+P(B|A_r)P(A_r)}{P(A_1)+\cdots+P(A_r)}$

$= P(B \mid A_1)P(A_1 \mid A_1 \cup \cdots \cup A_r) + \cdots + P(B \mid A_r)P(A_r \mid A_1 \cup \cdots \cup A_r)$

Es decir, dada la ocurrencia del evento $A_1 \cup \cdots \cup A_r$, cambiando las probabilidades $P(A_i)$ por $P(A_i \mid A_1 \cup \cdots \cup A_r)$, la probabilidad del evento B se calcula directamente usando la regla de la probabilidad total. Obsérvese que este resultado va de acuerdo con la idea de que dada la ocurrencia de un evento A, éste se vuelve el evento seguro.

En realidad el resultado anterior se puede ver como caso particular de la siguiente generalización de la regla de la probabilidad total.

PROPOSICIÓN 4.8 (**Regla generalizada de la probabilidad total**). *Sean A, B y $A_1, \ldots, A_n$ eventos relativos a un experimento $\mathcal{E}$ y supongamos que los eventos $A_1, \ldots, A_n$ forman una partición del espacio muestral y que $P(A) > 0$. Entonces:*

$$P(B \mid A) = \sum_{k=1}^{n} P(B \mid A \cap A_k)P(A_k \mid A)$$

Demostración

$P(B \mid A) = \frac{P(B\cap A)}{P(A)} = \frac{P[(B\cap A\cap A_1)\cup\ldots\cup(B\cap A\cap A_n)]}{P(A)} = \frac{P(B\cap A\cap A_1)+\ldots+P(B\cap A\cap A_n)}{P(A)}$

$= \frac{P(B|A\cap A_1)P(A\cap A_1)+\ldots+P(B|A\cap A_n)P(A\cap A_n)}{P(A)}$

$= P(B \mid A \cap A_1)\frac{P(A\cap A_1)}{P(A)} + \ldots + P(B \mid A \cap A_n)\frac{P(A\cap A_n)}{P\{A)}$

$= P(B \mid A \cap A_1)P(A_1 \mid A) + \ldots + P(B \mid A \cap A_n)P(A_n \mid A)$

■

Ejemplo 4.9 (**Ley de la sucesión de Laplace**). *Supongamos que tenemos $N+1$ urnas, cada una de las cuales contiene N bolas; supongamos además que la urna número k ($k \in \{0, \ldots, N\}$) contiene k bolas rojas y $N-k$ bolas blancas. Un experimento aleatorio consiste en seleccionar al azar una urna y, después de eso, en elegir, al azar y con reemplazo, $n+1$ bolas de la urna seleccionada. Si en las primeras n extracciones se han obtenido n bolas rojas, ¿cuál es la probabilidad de que la siguiente también resulte bola roja?*

Solución

Consideremos los siguientes eventos:

A_k*: la urna elegida es la k-ésima.*

B_i*: en cada una de las primeras i extracciones se obtiene bola roja.*

B*: en la última extracción se obtiene bola roja.*

Entonces:

$P(B_n) = \sum_{k=0}^{N} P(B_n \mid A_k)P(A_k) = \sum_{k=0}^{N} \left(\frac{k}{N}\right)^n \frac{1}{N+1} = \frac{1}{N+1} \sum_{k=0}^{N} \left(\frac{k}{N}\right)^n$

Los números $\frac{0}{N}, \frac{1}{N}, \ldots, \frac{N}{N}$ representan una partición del intervalo $[0,1]$, así que, cuando N es grande, la suma $\sum_{k=0}^{N} \left(\frac{k}{N}\right)^n \frac{1}{N}$ se puede aproximar por la integral entre 0 y 1 de la función x^n. Es decir, $\sum_{k=0}^{N} \left(\frac{k}{N}\right)^n \approx N \int_0^1 x^n dx = \frac{N}{n+1}$, de lo cual se obtiene:

$P(B_n) \approx \frac{N}{N+1} \frac{1}{n+1} \approx \frac{1}{n+1}$

De manera análoga, se tiene:

$P(B \cap B_n) = P(B_{n+1}) = \sum_{k=0}^{N} P(B_{n+1} \mid A_k)P(A_k) = \sum_{k=0}^{N} \left(\frac{k}{N}\right)^{n+1} \frac{1}{N+1}$

$= \frac{1}{N+1} \sum_{k=0}^{N} \left(\frac{k}{N}\right)^{n+1} \approx \frac{1}{N+1} \frac{N}{n+2} = \frac{N}{N+1} \frac{1}{n+2} \approx \frac{1}{n+2}$

Por lo tanto, tenemos:

$P(B \mid B_n) = \frac{P(B \cap B_n)}{P(B_n)} = \frac{\sum_{k=0}^{N} \left(\frac{k}{N}\right)^{n+1}}{\sum_{k=0}^{N} \left(\frac{k}{N}\right)^{n}} \approx \frac{n+1}{n+2}$

Es decir, mientras más grande sea la sucesión de bolas rojas obtenidas, la probabilidad de que la siguiente sea también roja se acercará más a 1.

El resultado es aparentemente paradójico pues pudiera pensarse que como las bolas se extraen al azar y con reemplazo de una urna fija, entonces el resultado de las primeras n extracciones no debería influir en la extracción siguiente. Sin embargo, si bien la *urna es fija una vez elegida, se trata de una urna elegida al azar de entre varias y es de esperarse que mientras más bolas rojas tenga la urna elegida más probable será que las primeras n bolas extraídas sean rojas y también más probable será que la última sea también roja. Esto es, intuitivamente puede pensarse que si en las primeras n extracciones se han obtenido n bolas rojas, entonces lo más probable es que se haya elegido una urna con muchas bolas rojas, de aquí que lo más probable es que en la última extracción también resulte bola roja. De una manera más precisa, utilizando la regla generalizada de la probabilidad total (véase proposición 4.8), tenemos:*

$P(B \mid B_n) = \sum_{k=0}^{N} P(B \mid B_n \cap A_k)P(A_k \mid B_n)$

Además, $P(B \mid B_n \cap A_k) = P(B \mid A_k)$, pues dado que la urna elegida es la k-ésima, la probabilidad de B no se altera por el resultado de las primeras n extracciones. Por lo tanto, se obtiene:

$P(B \mid B_n) = \sum_{k=0}^{N} P(B \mid A_k)P(A_k \mid B_n)$

Observemos entonces que, dado que el evento B_n ocurre, la situación que tenemos es equivalente a considerar un experimento consistente en elegir una de $N + 1$ posibles urnas, cada una de las cuales tiene una probabilidad distinta de ser elegida, dada por $P(A_k \mid B_n)$, respectivamente; después de elegir la urna se extrae una bola de la urna elegida y nos preguntamos entonces por la probabilidad de que la bola extraída sea roja. Si las urnas con mayor número de bolas rojas tienen mayor probabilidad de ser elegidas, entonces la probabilidad de extraer una bola roja será grande. Efectivamente, este es el caso, pues tenemos:

$P(A_k \mid B_n) = \frac{P(A_k \cap B_n)}{P(B_n)} = \frac{P(B_n|A_k)P(A_k)}{P(B_n)} = \frac{P(B_n|A_k)P(A_k)}{\sum_{j=)}^{N} P(B_n|A_j)P(A_j)}$

$= \frac{\left(\frac{k}{N}\right)^n \frac{1}{N+1}}{\sum_{j=0}^{N}\left(\frac{j}{N}\right)^n \frac{1}{N+1}} = \frac{\left(\frac{k}{N}\right)^n}{\sum_{j=0}^{N}\left(\frac{j}{N}\right)^n} = \frac{k^n}{\sum_{j=0}^{N} j^n}$

Así que mientras más grande sea k, más grande será $P(A_k \mid B_n)$.

Ejemplo 4.10 (**Aplicación a la Genética**). *En la reproducción de animales y plantas, algunas características de los individuos se transmiten a los descendientes; estas características son llamadas **caracteres hereditarios**.*

*Los caracteres hereditarios de un individuo están condicionados por el material genético transmitido por los padres. Éste está contenido en los **cromosomas** de los individuos de la especie y, en las especies heterosexuales, se transmite cuando una célula reproductora de un macho fecunda una célula reproductora de una hembra, formando una sola célula, en la cual está presente el material genético de cada una de las células reproductoras que la originaron.*

El número de cromosomas de cada célula reproductora es una constante N *para cada especie; así, por ejemplo, las células reproductoras del ser humano contienen* 23 *cromosomas. Al unirse una célula reproductora masculina con una femenina, se forman células con* N *pares de cromosomas, los* N *de cada una de ellas. Los cromosomas de cada par tienen la misma estructura y se llaman* ***cromosomas homólogos****.*

Cada cromosoma está compuesto de unidades de material genético, llamados ***genes****. En general, con cada gen de un cromosoma se encuentra asociado un gen del cromosoma homólogo. Cada par de genes homólogos, o un grupo de pares de ellos, determina un carácter hereditario, tal como el color de las flores de una planta, el color de los ojos de un animal, el tamaño de la nariz o la estatura de un ser humano.*

Genes del mismo tipo no producen siempre el mismo efecto; por ejemplo, en dos plantas de la misma especie, el color de las flores está determinado por el mismo tipo de gen, sin embargo, en una el color puede ser blanco y en otra rojo. En otras palabras, un mismo tipo de gen se puede presentar de diferentes formas, las cuales representaremos mediante letras mayúsculas $A, B, C, \ldots$ *o minúsculas* $a, b, c, \ldots$. *El efecto que se produzca en un individuo dependerá de la forma en que se presente cada uno del par de genes homólogos correspondientes al carácter hereditario en consideración.*

Muchos genes se pueden presentar únicamente en una de dos posibles formas; tal es el caso de los genes que determinan el color de algunas flores o el sexo de un ser humano; sin embargo, otros genes se pueden presentar en una de tres o más formas posibles; tal es el caso de los genes que determinan el tipo de sangre en un ser humano, los cuales se pueden presentar en una de tres posibles formas.

En lo que sigue consideraremos únicamente caracteres hereditarios determinados por genes que pueden presentarse únicamente en una de dos posibles formas, A *y* a*. El efecto sobre un individuo dependerá de las formas en que se presenten los genes en los correspondientes pares de cromosomas homólogos. Son tres las posibilidades,* AA*-*Aa*-*aa*, a cada una de las cuales llamaremos un* ***genotipo****. Diremos que un individuo con genotipo* AA *o* aa *es* ***puro****, mientras que uno con genotipo* Aa *será llamado* ***híbrido****.*

Una célula reproductora contiene únicamente uno de los cromosomas de cada par de cromosomas homólogos, de manera que una de estas células contiene únicamente uno de los dos genes presentes en el par de cromosomas homólogos. Este gen, presente en una de las formas A *o* a *en las células reproductoras, se transmite a los descendientes. El genotipo de un descendiente dependerá entonces del genotipo de cada uno de los padres y de la forma,* A *o* a*, presente en el genotipo que se transmita.*

Cuando el genotipo de un padre es AA *o* aa*, no existe ningún problema para determinar cual forma,* A *o* a*, se transmite pues únicamente hay una posibilidad. Así, si un macho de genotipo* AA *se cruza con una hembra de genotipo* aa*, los descendientes* *tendrán genotipo Aa. El problema se presenta cuando uno de los padres tiene genotipo Aa, pues en ese caso hay dos posibilidades para la forma que se transmite.*

En algunos casos la presencia de un genotipo puede no ser evidente pues una de las formas del gen puede ser ***dominante****, lo cual significa que se manifestará siempre que esté presente. Por ejemplo, al cruzar dos líneas de chícharos, una que produce flores blancas y otra flores de color, se observa que todos los descendientes producen flores de color; sin embargo, al cruzar los descendientes, se observa que los nuevos descendientes producen tanto flores de color como flores blancas. Este resultado se explica si suponemos que la forma A es dominante y, siempre que se presente en el genotipo de un individuo, tendrá como efecto el que se produzcan flores de color. La otra forma a, que llamaremos* ***recesiva****, tendrá el efecto de que se produzcan flores blancas únicamente cuando el genotipo sea aa. Así, en el ejemplo de los chícharos, se comienza cruzando individuos de genotipo AA con individuos de genotipo aa, dando como resultado descendientes de genotipo Aa, los cuales, por estar presente A en el genotipo, producirán flores de color todos ellos. Al cruzar los descendientes, se producirán individuos con genotipo AA, Aa o aa, es decir de flores blancas o de color.*

Gregor Johann Mendel (1822-1884), experimentando con la reproducción de semillas de chícharo, descubrió que los resultados experimentales, respecto a un carácter hereditario, se explican si se supone que, al cruzarse dos individuos de una población, cada uno de ellos puede transmitir a los descendientes una de las dos formas presentes en su genotipo, siendo iguales las probabilidades de transmitir una u otra de esas formas y siendo independientes una de otra la transmisión que realiza cada individuo de la pareja.

En este ejemplo vamos a utilizar este modelo para estudiar la manera en que evoluciona la proporción de cada genotipo, de un determinado carácter hereditario, en una población heterosexual. Asumiremos que el carácter hereditario en consideración está determinado por un tipo de gen que se puede presentar únicamente en dos posibles formas, que denotaremos por A y a, dando lugar entonces a los genotipos AA, Aa y aa. Asumiremos también que los individuos de la población se reproducen con cruzamientos al azar entre dos individuos de sexo opuesto presentes en la población. La población inicial Π_0 *da origen, mediante reproducción, a una población* Π_1*, la cual a su vez origina una población* Π_2*, etc.*

Para cada $n \in \{0, 1, \ldots\}$*, sean* u_n, v_n, w_n *(resp.* u_n', v_n', w_n'*) las frecuencias relativas de los genotipos AA, Aa y aa, respectivamente, entre los machos (resp. hembras) de la población* Π_n*. Además, denotaremos por* p_n *y* q_n *(resp.* p_n' *y* q_n'*) a las probabilidades de que en un apareamiento de la población* Π_n*, el macho (resp. hembra) aporte las* *formas A y a del gen, respectivamente. Evidentemente se tiene* $u_n + v_n + w_n = 1$, $u'_n + v'_n + w'_n = 1$, $p_n + q_n = 1$ *y* $p'_n + q'_n = 1$.

Por la regla de la probabilidad total se tiene:

$p_n = u_n + \frac{1}{2}v_n$

$q_n = w_n + \frac{1}{2}v_n$

$p'_n = u'_n + \frac{1}{2}v'_n$

$q'_n = w'_n + \frac{1}{2}v'_n$

Así que, por la regla del producto y la propiedad de la aditividad finita:

$u'_{n+1} = u_{n+1} = p_n p'_n = \left(u_n + \frac{1}{2}v_n\right)\left(u'_n + \frac{1}{2}v'_n\right)$

$v'_{n+1} = v_{n+1} = p_n q'_n + p'_n q_n = \left(u_n + \frac{1}{2}v_n\right)\left(w'_n + \frac{1}{2}v'_n\right) + \left(u'_n + \frac{1}{2}v'_n\right)\left(w_n + \frac{1}{2}v_n\right)$

$w'_{n+1} = w_{n+1} = q_n q'_n = \left(w_n + \frac{1}{2}v_n\right)\left(w'_n + \frac{1}{2}v'_n\right)$

De manera que, para $n \geq 1$, *se tiene:*

$u'_{n+1} = u_{n+1} = \left(u_n + \frac{1}{2}v_n\right)^2$

$v'_{n+1} = v_{n+1} = 2\left(u_n + \frac{1}{2}v_n\right)\left(w_n + \frac{1}{2}v_n\right)$

$w'_{n+1} = w_{n+1} = \left(w_n + \frac{1}{2}v_n\right)^2$

Por lo tanto, para $n \geq 2$, *se tiene:*

$u'_{n+1} = u_{n+1} = \left(u_n + \frac{1}{2}v_n\right)^2$

$= \left[\left(u_{n-1} + \frac{1}{2}v_{n-1}\right)^2 + \left(u_{n-1} + \frac{1}{2}v_{n-1}\right)\left(w_{n-1} + \frac{1}{2}v_{n-1}\right)\right]^2$

$= \left(u_{n-1} + \frac{1}{2}v_{n-1}\right)^2 \left[\left(u_{n-1} + \frac{1}{2}v_{n-1}\right) + \left(w_{n-1} + \frac{1}{2}v_{n-1}\right)\right]^2$

$= \left(u_{n-1} + \frac{1}{2}v_{n-1}\right)^2 \left[u_{n-1} + v_{n-1} + w_{n-1}\right]^2 = \left(u_{n-1} + \frac{1}{2}v_{n-1}\right)^2 = u_n$

$w'_{n+1} = w_{n+1} = \left(w_n + \frac{1}{2}v_n\right)^2$

$= \left[\left(w_{n-1} + \frac{1}{2}v_{n-1}\right)^2 + \left(u_{n-1} + \frac{1}{2}v_{n-1}\right)\left(w_{n-1} + \frac{1}{2}v_{n-1}\right)\right]^2$

$= \left(w_{n-1} + \frac{1}{2}v_{n-1}\right)^2 \left[\left(w_{n-1} + \frac{1}{2}v_{n-1}\right) + \left(u_{n-1} + \frac{1}{2}v_{n-1}\right)\right]^2$

$= \left(w_{n-1} + \frac{1}{2}v_{n-1}\right)^2 [w_{n-1} + v_{n-1} + u_{n-1}]^2 = \left(w_{n-1} + \frac{1}{2}v_{n-1}\right)^2 = w_n$

$v'_{n+1} = v_{n+1} = 1 - u_{n+1} - w_{n+1} = 1 - u_n - w_n = v_n$

*Se obtiene así lo que se conoce como la **ley de Hardy**:* a partir de la segunda generación, la frecuencia de cada uno de los genotipos se estabiliza.

El sexo de un individuo es determinado por un tipo de gen que se presenta en dos formas, denotadas usualmente por X y Y, una de las cuales es dominante, usualmente la que es denotada por Y. En la mayoría de los casos, siempre que se presenta la forma Y, el individuo es de sexo masculino, tal es el caso de la especie humana, de la mosca de la fruta y de muchas otras especies; pero, en aves, mariposas y otras especies, la presencia de la forma Y en un individuo determina que su sexo sea femenino.

Existen algunos caracteres determinados por genes que se presentan únicamente en los cromosomas que contienen la forma X. Ese tipo de genes se presenta en general bajo dos formas, A y a, de las cuales una es dominante, digamos A. Tal es el caso de los genes que determinan el color de los ojos de la mosca de la fruta, siendo rojo en todos los individuos que tengan en su genotipo por lo menos una forma A y blanco en los otros casos.

*En todos los casos en que un carácter hereditario esté determinado por genes que se presentan únicamente en los cromosomas que contienen la forma X, diremos que el carácter, o el gen, está **ligado al sexo**. Cuando un gen ligado al sexo se presenta únicamente en una de dos posibles formas A y a, de las cuales la primera es dominante y la segunda recesiva, diremos que un individuo es **recesivo** cuando se encuentre ausente la forma A en su genotipo.*

Algunos genes ligados al sexo causan defectos en los individuos recesivos, tal es el caso de los genes que producen el daltonismo y la hemofilia en los seres humanos.

Consideraremos ahora una población en la cual la forma XY determina sexo masculino y la forma XX sexo femenino. Consideraremos además un gen ligado al sexo de esa población, que se presenta en forma dominante y recesiva, y asumiremos que los individuos de la población se reproducen con cruzamientos al azar entre dos individuos de sexo opuesto presentes en la población. La población inicial Π_0 da origen, mediante reproducción, a una población Π_1, la cual a su vez origina una población Π_2, etc.

Para cada $n \in \{0, 1, \ldots\}$, sean u_n, v_n, w_n las frecuencias relativas de los genotipos AA, Aa y aa, respectivamente, entre las hembras de la población Π_n y r_n, s_n las frecuencias relativas de los genotipos A y a, respectivamente, entre los machos. Además, denotaremos por p_n y q_n a las probabilidades de que en un apareamiento de la población

Π_n, la hembra aporte las formas A y a del gen, respectivamente. Evidentemente, se tiene $u_n + v_n + w_n = 1$, $r_n + s_n = 1$ y $p_n + q_n = 1$.

Por la regla de la probabilidad total se tiene:

$p_n = u_n + \frac{1}{2}v_n$

$q_n = w_n + \frac{1}{2}v_n$

Mientras que, por la regla del producto y la propiedad de la aditividad finita, se tiene:

$u_{n+1} = p_n r_n$

$v_{n+1} = p_n s_n + q_n r_n$

$w_{n+1} = q_n s_n$

Finalmente, se tiene también:

$r_{n+1} = p_n$

$s_{n+1} = q_n$

Así que:

$p_{n+1} = u_{n+1} + \frac{1}{2}v_{n+1} = p_n r_n + \frac{1}{2}(p_n s_n + q_n r_n) = \frac{1}{2}(p_n r_n + p_n s_n) + \frac{1}{2}(p_n r_n + q_n r_n) = \frac{1}{2}(p_n + r_n)$

De manera que:

$p_1 - p_0 = \frac{1}{2}(p_0 + r_0) - p_0 = -\frac{1}{2}(p_0 - r_0)$

Y, para $n \geq 1$:

$p_{n+1} - p_n = \frac{1}{2}(p_n + r_n) - p_n = \frac{1}{2}(p_n + p_{n-1}) - p_n = -\frac{1}{2}(p_n - p_{n-1})$

Por lo tanto, para $n \geq 1$, se tiene:

$p_n - p_{n-1} = (-1)^n \frac{1}{2^n}(p_0 - r_0)$

$p_n = \sum_{k=1}^{n}(p_k - p_{k-1}) + (p_0 - r_0) + r_0 = r_0 + \sum_{k=0}^{n}(-1)^k \frac{1}{2^k}(p_0 - r_0)$

$= r_0 + \frac{1-(-1)^{n+1}\frac{1}{2^{n+1}}}{1+\frac{1}{2}}(p_0 - r_0) = r_0 + \left[\frac{2}{3} + (-1)^n \frac{1}{3\cdot 2^n}\right](p_0 - r_0)$

$= \frac{2}{3}p_0 + \frac{1}{3}r_0 + (-1)^n \frac{1}{3\cdot 2^n}(p_0 - r_0)$

$q_n = 1 - p_n = 1 - \left[\frac{2}{3}p_0 + \frac{1}{3}r_0 + (-1)^n \frac{1}{3 \cdot 2^n}(p_0 - r_0)\right]$

$= 1 - \left[1 - \frac{2}{3}q_0 - \frac{1}{3}s_0 - (-1)^n \frac{1}{3 \cdot 2^n}(q_0 - s_0)\right]$

$= \frac{2}{3}q_0 + \frac{1}{3}s_0 + (-1)^n \frac{1}{3 \cdot 2^n}(q_0 - s_0)$

$r_n = p_{n-1} = \frac{2}{3}p_0 + \frac{1}{3}r_0 + (-1)^{n-1} \frac{1}{3 \cdot 2^{n-1}}(p_0 - r_0)$

$s_n = q_{n-1} = \frac{2}{3}q_0 + \frac{1}{3}s_0 + (-1)^{n-1} \frac{1}{3 \cdot 2^{n-1}}(q_0 - s_0)$

Si n *es grande, se tiene entonces:*

$p_n \approx r_n \approx \frac{2}{3}p_0 + \frac{1}{3}r_0 = 1 - \left(\frac{2}{3}q_0 + \frac{1}{3}s_0\right)$

$q_n \approx s_n \approx \frac{2}{3}q_0 + \frac{1}{3}s_0$

Definamos $\alpha = \frac{2}{3}q_0 + \frac{1}{3}s$*, entonces se tiene:*

$p_n \approx r_n \approx 1 - \alpha$

$q_n \approx s_n \approx \alpha$

$u_n = p_{n-1}r_{n-1} \approx (1 - \alpha)^2$

$v_n = p_{n-1}s_{n-1} + q_{n-1}r_{n-1} \approx 2\alpha(1 - \alpha)$

$w_n = q_{n-1}s_{n-1} \approx \alpha^2$

Del resultado anterior, se sigue, en particular, que si, a la larga, α *es la frecuencia relativa de machos recesivos, entonces la frecuencia de hembras recesivas será* α^2*. Así, por ejemplo, si* 1 *de cada* 100 *machos serán recesivos, entonces* 1 *de cada* $10,000$ *hembras serán recesivas. Esto explica, por ejemplo, el que la hemofilia y el daltonismo sean más frecuentes entre los hombres que entre las mujeres.*

4.2. Regla de Bayes

La probabilidad condicional surge de problemas del siguiente tipo: se realizan en forma sucesiva dos experimentos aleatorios tales que, una vez hecho el primero, las probabilidades de los posibles resultados del segundo dependen del resultado de aquél; entonces, si A y B son dos eventos relativos al primero y segundo experimento, respectivamente, nos interesa conocer la probabilidad de que ocurra el evento B dado que ocurre el evento A. Sin embargo, la definición de probabilidad condicional es general y, en la misma situación descrita, tiene sentido calcular la probabilidad de ocurrencia del evento A dado que ocurre el evento B. A este tipo de probabilidades se les conoce como probabilidades de causas pues si pensamos en los eventos A y B como relativos a un primero y segundo experimentos, respectivamente, podemos pensar también a los resultados del primer experimento como las "causas" de los resultados del segundo; evidentemente el nombre "causas" expresa únicamente que hay un orden en la realización de los experimentos y no una relación real de causa-efecto.

En general, en una situación como la descrita, es más simple calcular $P(B \mid A)$ que $P(A \mid B)$, por tal motivo el cálculo de esta última se realiza utilizando la siguiente fórmula:

$$P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$$

relación que es conocida como **fórmula de Bayes** por haber sido Thomas Bayes el primero en usarla explícitamente, aunque resulta inmediatamente de la regla del producto y ésta ya había sido demostrada en su forma general por Abraham de Moivre.

Consideremos ahora un experimento aleatorio en el cual necesariamente ocurre alguno de n eventos mutuamente excluyentes $A_1, \ldots, A_n$. Sea B un evento relativo al experimento; usando la regla de la probabilidad total, la fórmula de Bayes puede ser escrita de la siguiente manera:

$$P(A_i \mid B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1)+\cdots+P(B|A_n)P(A_n)}$$

EJEMPLO 4.11. *Una urna contiene r bolas rojas y b bolas blancas. Se sacan sucesivamente y sin reemplazo, dos bolas de la urna. Calcule la probabilidad de que la primera bola sea roja bajo la hipótesis de que la segunda lo es.*

Solución

Consideremos los eventos siguientes:

A_1*: la primera bola es roja.*

A_2*: la primera bola es blanca.*

B*: la segunda bola es roja.*

Entonces:

$$P(A_1 \mid B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)} = \frac{r-1}{r+b-1}$$

Obsérvese que, en este caso, $P(A_1 \mid B) = P(B \mid A_1)$, lo cual es explicable pues en realidad la segunda extracción puede ser considerada como la primera y viceversa.

EJEMPLO 4.12. *Una urna contiene $2N$ bolas numeradas del 1 al $2N$. Un experimento consiste en elegir, al azar y sin reemplazo, dos bolas de la urna, consecutivamente. Calcule la probabilidad de que en la primera elección resulte un número par bajo la hipótesis de que la suma es par.*

Solución

Consideremos los eventos siguientes:

B: la suma de los dos números elegidos es par.

A_1 : el número de la primera elección es par.

A_2 : el número de la primera elección es impar.

Entonces:

$P(A_1 \mid B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)} = \frac{1}{2}$

EJEMPLO 4.13. *Supongamos que tenemos N urnas, cada una de las cuales contiene N bolas; supongamos además que la urna número i ($i \in \{1, \ldots, N\}$) contiene i bolas rojas y $N - i$ bolas blancas. Un experimento está compuesto de dos partes, en la primera se selecciona al azar una urna, fijando así la probabilidad p de que, al seleccionar una bola de ella, se obtenga una roja. En la segunda parte se seleccionan, al azar y con reemplazo, n bolas de la urna seleccionada. Sabiendo que la muestra de n bolas contiene k bolas rojas, ¿cuál es la probabilidad de que el valor de p, que se fija en la primera parte del experimento, esté comprendido entre dos valores α y β, en donde $0 \leq \alpha \leq \beta \leq 1$?*

Solución

Consideremos los siguientes eventos:

A_i: p toma el valor $p_i = \frac{i}{N}$, para $i \in \{1, \ldots, N\}$.

A: p está comprendida entre α y β.

B_k: la muestra de n bolas contiene exactamente k rojas.

Se tiene:

$P(A \mid B_k) = \frac{P(A \cap B_k)}{P(B_k)} = \frac{\sum_{i=1}^{N} P(A \cap B_k \cap A_i)}{\sum_{i=1}^{N} P(B_k|_i)P(A_i)} = \frac{\sum_{\{i:\alpha \leq \frac{i}{N} \leq \beta\}} P(B \cap A_i)}{\sum_{i=1}^{N} P(B_k|A_i)P(A_i)}$

$= \frac{\sum_{\{i:\alpha \leq \frac{i}{N} \leq \beta\}} P(B_k|A_i)P(A_i)}{\sum_{i=1}^{N} P(B_k|A_i)P(A_i)} = \frac{\sum_{\{i:\alpha \leq \frac{k}{N} \leq \beta\}} \binom{n}{k} p_i^k (1-p_i)^{n-k} \frac{1}{N}}{\sum_{i=1}^{N} \binom{n}{k} p_i^k (1-p_i)^{n-k} \frac{1}{N}}$

$$= \frac{\sum_{\{i:\alpha \leq \frac{i}{N} \leq \beta\}} \binom{n}{k}\left(\frac{i}{N}\right)^k \left(1-\frac{i}{N}\right)^{n-k} \frac{1}{N}}{\sum_{i=1}^{N} \binom{n}{k}\left(\frac{i}{N}\right)^k \left(1-\frac{i}{N}\right)^{n-k} \frac{1}{N}}$$

Con N fija, cada suma de este cociente es una suma de Riemann correspondiente a la función $\binom{n}{k}x^k(1-x)^{n-k}$ y a una partición del intervalo $[\alpha, \beta]$ y $[0,1]$, respectivamente, en partes iguales. Por lo tanto, si N es grande, tendremos:

$$P(A \mid B_k) \approx \frac{\int_\alpha^\beta \binom{n}{k}x^k(1-x)^{n-k}dx}{\int_0^1 \binom{n}{k}x^k(1-x)^{n-k}dx} = (n+1)\binom{n}{k} \int_\alpha^\beta x^k(1-x)^{n-k}dx$$

Este resultado se debe a Bayes, quien se planteaba el problema de determinar la probabilidad de ocurrencia de un evento a partir del conocimiento del resultado de n repeticiones del experimento con respecto al cual está definido dicho evento. Si p es la probabilidad buscada, Bayes encuentra que, dado que en las n repeticiones del experimento el evento ocurre k veces, la probabilidad de que p esté comprendida entre dos números α y β está dada por la fórmula de arriba.

La gráfica de la función $x \mapsto \binom{n}{k}x^k(1-x)^{n-k}$ es como sigue:

$n = 20,\ k = 15$

Obsérvese que el valor máximo se alcanza en $x_0 = \frac{k}{n}$; esto significa que, dado que el evento ocurre k veces en n repeticiones del experimento, es más probable que p esté comprendida en un intervalo que contenga al punto $\frac{k}{n}$.

El tomar $p_i = \frac{i}{n}$ al inicio del experimento lo justifica Bayes por el hecho de que, de esa manera, la probabilidad de obtener exactamente k bolas rojas en la muestra de n es la misma cualquiera que sea k, en efecto:

$$P(B_k) = \sum_{i=1}^{N} \binom{n}{k}(\tfrac{i}{N})^k(1-\tfrac{i}{N})^{n-k}\tfrac{1}{N} \approx \int_0^1 \binom{n}{k}x^k(1-x)^{n-k}dx = \tfrac{1}{n+1}$$

En su forma original, Bayes no parte de las probabilidades $p_i = \frac{i}{n}$ *sino de un modelo continuo. Bajo esa forma, este problema volverá a tratarse en la sección de distribuciones condicionales, en el segundo volumen de este libro.*

EJEMPLO 4.14. *Una moneda balanceada es lanzada* N *veces. En cada lanzamiento se coloca en una urna una bola blanca, cuando el resultado es cara y una bola roja, cuando el resultado es cruz. En seguida se sacan con reemplazo* n *bolas de la urna. Calcule la probabilidad de que la proporción de bolas blancas en la urna esté comprendida entre dos números* α *y* β *bajo la hipótesis de que de las* n *bolas extraídas,* k *son blancas.*

Solución

Consideremos los siguientes eventos:

A_i *: la urna contiene* i *bolas blancas.*

B_k*: de las* n *bolas extraídas,* k *son blancas.*

A: la proporción de bolas blancas en la urna esta comprendida entre α *y* β*.*

Entonces:

$$P(A \mid B_k) = \frac{P(A \cap B_k)}{P(B_k)} = \frac{\sum_{i=1}^{N} P(A \cap B_k \cap A_i)}{\sum_{i=1}^{N} P(B_k|_i)P(A_i)} = \frac{\sum_{\{i:\alpha \leq \frac{i}{N} \leq \beta\}} P(B_k \cap A_i)}{\sum_{i=1}^{N} P(B_k|A_i)P(A_i)}$$

$$= \frac{\sum_{\{i:\alpha \leq \frac{i}{N} \leq \beta\}} P(B_k|A_i)P(A_i)}{\sum_{i=1}^{N} P(B_k|A_i)P(A_i)} = \frac{\sum_{\{i:\alpha \leq \frac{i}{N} \leq \beta\}} \binom{n}{k}(\frac{i}{N})^k(1-\frac{i}{N})^{n-k}\binom{N}{i}\frac{1}{2^N}}{\sum_{i=1}^{N} \binom{n}{k}(\frac{i}{N})^k(1-\frac{i}{N})^{n-k}\binom{N}{i}\frac{1}{2^N}}$$

$$= \frac{\sum_{\{i:\alpha \leq \frac{i}{N} \leq \beta\}} \binom{N}{i}\left(\frac{i}{N}\right)^k\left(1-\frac{i}{N}\right)^{n-k}}{\sum_{i=1}^{N} \binom{N}{i}\left(\frac{i}{N}\right)^k\left(1-\frac{i}{N}\right)^{n-k}}$$

▲

Como se ilustra con el siguiente ejemplo, la regla de Bayes puede ser útil aún cuando no se trate de experimentos que constan de varias partes.

EJEMPLO 4.15. *Cuando se realiza una prueba para detectar la presencia de una cierta enfermedad, se dice que se obtiene un falso negativo cuando la prueba resulta negativa para una persona que tiene la enfermedad, mientras que se dice que se obtiene un falso positivo cuando la prueba resulta positiva para una persona que no la tiene. Una buena prueba es entonces aquella para la cual la probabilidad de un falso negativo y de un falso positivo sean pequeñas. Supongamos que una prueba de sangre de una cierta enfermedad contagiosa es tal que las probabilidades de un falso negativo y de un falso positivo son ambas iguales a* 0.05*. Supongamos además que la probabilidad de que una persona elegida al azar esté contagiada es igual a* p*. Se elige una persona al azar y se* *le practica la prueba, resultando positiva, ¿cuál es la probabilidad de que la persona esté contagiada?*

Solución

Definamos los eventos

A*: la persona seleccionada está contagiada.*

B*: la prueba resulta positiva.*

Se sabe entonces que $P(B^c \mid A) = 0.05$, $P(B \mid A^c) = 0.05$ *y* $P(A) = p$.

Por lo tanto:

$P(A \mid B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)} = \frac{0.95p}{0.95p+0.05(1-p)} = \frac{0.95p}{0.90p+0.05}$

La gráfica de la función $f(p) = \frac{0.95p}{0.90p+0.05}$ *se muestra a continuación.*

Obsérvese que cuando p es pequeña, dado que la prueba resulta positiva, la probabilidad de que la persona esté contagiada no es muy grande. Por ejemplo, $f(0.005) = 0.087$, *de manera que, en ese caso, menos del* 9% *de la personas para las cuales la prueba resulta positiva, están contagiadas. La probabilidad de contagio es alta únicamente para valores de p relativamente altos, por ejemplo* $f(0.3) = 0.89063$, *de manera que, en ese caso, más del* 89% *de las personas para las cuales la prueba resulta positiva, están contagiadas.*

EJERCICIOS

EJERCICIO 4.1. *Una urna contiene* 2 *bolas blancas y* 4 *bolas rojas, una segunda urna contiene* 3 *bolas blancas y* 2 *bolas rojas y una tercera urna contiene* 4 *bolas blancas y* 5 *bolas rojas. Se elige* 1 *bola al azar de la primera urna y se transfiere a la segunda; después de esto, se selecciona al azar* 1 *bola de la segunda urna y se transfiere a la tercera; finalmente se selecciona al azar una bola de la tercera urna. Calcule la probabilidad de que la bola que se obtiene finalmente sea roja.*

EJERCICIO 4.2. *Una urna contiene* 2 *bolas blancas y* 8 *bolas rojas, una segunda urna contiene* 4 *bolas blancas y* 4 *bolas rojas. Se eligen, al azar y sin reemplazo,* 2 *bolas de la primera urna y se transfieren a la segunda; después de esto, se selecciona al azar* 1 *bola de la segunda urna. Calcule la probabilidad de que la bola que se obtiene finalmente sea roja.*

EJERCICIO 4.3. *De una urna que contiene* 6 *bolas blancas y* 4 *bolas rojas se seleccionan, sin reemplazo,* 6 *bolas al azar y se transfieren a una segunda urna, la cual se encuentra vacía. Inmediatamente después, se seleccionan, al azar y sin reemplazo,* 4 *bolas de la segunda urna. Calcule la probabilidad de que entre las* 4 *bolas seleccionadas de la segunda urna haya exactamente* 2 *blancas.*

EJERCICIO 4.4. *De una urna que contiene* 6 *bolas blancas y* 4 *bolas rojas se seleccionan, sin reemplazo,* 4 *bolas al azar y se transfieren a una segunda urna, la cual se encuentra vacía. Inmediatamente después, se seleccionan, al azar y sin reemplazo, dos bolas de la segunda urna. Calcule la probabilidad de que las dos bolas seleccionadas de la segunda urna sean ambas blancas.*

EJERCICIO 4.5. *Una urna contiene* 8 *bolas blancas y* 2 *bolas rojas, una segunda urna contiene* 5 *bolas blancas y* 5 *bolas rojas y una tercera urna contiene* 3 *bolas blancas y* 5 *bolas rojas. Se elige* 1 *bola al azar de cada una de las dos primeras urnas y se transfieren a la tercera; después de esto, se seleccionan al azar* 2 *bolas de la tercera urna. Calcule la probabilidad de que las dos bolas seleccionadas de la tercera urna sean ambas blancas.*

EJERCICIO 4.6. *Una urna contiene* 4 *bolas rojas y* 6 *blancas; una segunda contiene* 6 *bolas rojas y* 2 *blancas y una tercera contiene* 3 *bolas rojas y* 5 *blancas. Un experimento aleatorio consiste en elegir al azar una de las urnas y, después de eso, en elegir al azar una bola de la urna elegida. Sabiendo que la bola seleccionada no proviene de la tercera urna, ¿cuál es la probabilidad de que sea roja?*

EJERCICIO 4.7. *Una urna contiene* r *bolas rojas y* b *bolas blancas. Se elige al azar una bola de la urna, después, sin reemplazar la primera, se elige otra y así sucesivamente, hasta sacar todas las bolas. Utilizando la regla de la probabilidad total y mediante un* *razonamiento de inducción, demuestre que la probabilidad de que en el j-ésimo paso se saque una bola roja es igual a $\frac{r}{r+b}$.*

EJERCICIO 4.8. *Consideremos una población formada por familias en las cuales hay a lo más 5 hijos, de tal manera que los porcentajes de familias con 0, 1, 2 ,3, 4 y 5 hijos están dados por 5%, 15%, 30%, 25%, 20% y 5%, respectivamente. Supongamos además que la probabilidad de que un hijo sea varón es igual a 0.5. Al seleccionar una familia al azar de esa población, ¿cuál es la probabilidad de que todos los hijos sean varones sabiendo que a) hay al menos un hijo varón? y b) la familia tiene por lo menos un hijo?*

EJERCICIO 4.9. *Se tienen 3 cartas; una de ellas es roja de ambos lados, la segunda es roja de un lado y blanca del otro y la tercera es blanca de ambos lados. Se elige al azar una de las cartas y se coloca sobre una mesa. Sabiendo que el color que muestra la carta seleccionada es rojo, ¿cuál es la probabilidad de que del otro lado también sea roja?*

EJERCICIO 4.10. *Se tienen 4 escritorios, cada uno con dos cajones. Los dos primeros escritorios contienen una moneda de oro en cada cajón, el tercero contiene una moneda de plata en cada cajón y el cuarto contiene una moneda de oro en un cajón y una de plata en el otro. Se elige un escritorio al azar, se abre un cajón al azar y después el otro. Si en el primer cajón se encuentra una moneda de oro, ¿cuál es la probabilidad de que el segundo cajón también contenga una moneda de oro?*

EJERCICIO 4.11. *La probabilidad de que un documento se encuentre en alguno de los 8 cajones de un escritorio es igual a p. Si el documento se encuentra en el escritorio, es igualmente probable que se encuentre en cualquiera de sus 8 cajones. Se busca el documento en 7 de los cajones y no se encuentra. Calcule la probabilidad de encontrarlo en el octavo cajón.*

EJERCICIO 4.12. *Una urna contiene 2 bolas blancas y 8 bolas rojas, una segunda urna contiene 4 bolas blancas y 4 bolas rojas. Se eligen, al azar y sin reemplazo, 2 bolas de la primera urna y se transfieren a la segunda; después de esto, se seleccionan, al azar y sin reemplazo, 3 bolas de la segunda urna. Calcule la probabilidad de que entre estas últimas haya bolas de los dos colores.*

EJERCICIO 4.13. *Una urna contiene 2 bolas blancas y 1 negra; otra urna contiene 1 bola blanca y 5 bolas negras. Una bola elegida al azar es transferida de la primera urna a la segunda y, después de ésto, se extrae al azar una bola de esta última. Calcule la probabilidad de que la bola que se transfiere sea negra bajo la hipótesis de que en la extracción resulta bola blanca.*

EJERCICIO 4.14. *Se selecciona al azar una bola de una urna, la cual contiene 3 bolas rojas y 6 blancas. Si la bola seleccionada es blanca, se devuelve a la urna, mientras que si es roja, no se devuelve. Inmediatamente después se elige al azar otra bola de* *la urna. Si se sabe que la segunda bola seleccionada es roja, ¿cuál es la probabilidad de que la primera también lo sea?*

Ejercicio 4.15. *Se selecciona al azar una bola de una urna, la cual contiene* 10 *bolas rojas y* 5 *bolas negras. Si la bola seleccionada es roja, se regresa a la urna, mientras que si es negra, la bola seleccionada se regresa a la urna junto con dos bolas negras más. a) Calcule la probabilidad de que una segunda bola que se seleccione al azar de la urna sea negra. b) Si al seleccionar una segunda bola al azar se obtiene una bola negra, ¿cuál es la probabilidad de que la primera bola que se selecciona también sea negra?*

Ejercicio 4.16. *De una urna que contiene* 3 *bolas rojas y* 6 *bolas blancas se selecciona al azar una bola y se regresa a la urna junto con* 3 *bolas más del mismo color; inmediatamente después se selecciona al azar otra bola de la urna. Sabiendo que la segunda bola seleccionada es roja, ¿cuál es la probabilidad de que la primera también lo sea?*

Ejercicio 4.17. *Se tienen dos urnas, la primera contiene* 3 *bolas blancas y* 4 *bolas rojas, mientras que la segunda contiene* 6 *bolas blancas y* 3 *bolas rojas. Un experimento aleatorio consiste en lanzar una moneda balanceada una vez; si se obtiene cara, se selecciona al azar una bola de la primera urna, si se obtiene cruz, se selecciona al azar una bola de la segunda urna. Sabiendo que finalmente se obtiene una bola blanca, ¿cuál es la probabilidad de que se obtenga cara al lanzar la moneda?*

Ejercicio 4.18. *Una moneda desbalanceada está hecha de tal forma que la probabilidad de obtener cara es el doble de la probabilidad de obtener cruz. Un experimento aleatorio consiste en lanzar dicha moneda una vez, si se obtiene cara, se selecciona al azar una bola de una urna que contiene* 9 *bolas blancas y* 3 *bolas rojas, mientras que si se obtiene cruz, se selecciona al azar una bola de otra urna que contiene* 4 *bolas blancas y* 8 *bolas rojas. a) ¿Cuál es la probabilidad de que se obtenga finalmente una bola roja? b) Supongamos que como resultado del experimento se obtiene finalmente una bola roja, ¿cuál es la probabilidad de que en la primera parte de éste se obtenga cruz al lanzar la moneda?*

Ejercicio 4.19. *En una cierta población, el* 70 % *de los individuos son hombres y el* 30 % *son mujeres. Se sabe que el* 40 % *de los hombres de esa población fuma, mientras que de las mujeres lo hace el* 60 %. *a) Si se seleccionada al azar una persona de dicha población, ¿cuál es la probabilidad de que esa persona fume? b) Si se observa que un individuo de dicha población está fumando, ¿cuál es la probabilidad de que esa persona sea una mujer?*

Ejercicio 4.20. *Una compañía de seguros estima que en una cierta población, el* 30 % *de los individuos son propensos a tener accidentes. Estima además que la probabilidad de que una persona propensa a tener dicho accidente lo tenga realmente, en un periodo de un año, es igual a* 0.4, *mientras que esta probabilidad se reduce a* 0.2 *para las* *personas no propensas. a) ¿Cuál es la probabilidad de que un nuevo asegurado tenga un accidente durante el primer año de la póliza?. Si un asegurado tiene un accidente durante el primer año de la póliza, ¿cuál es la probabilidad de que b) esa persona sea propensa a tener un accidente? y c) el asegurado tenga un accidente durante el segundo año de la póliza?*

Ejercicio 4.21. *Un estudiante presenta un examen de opción múltiple en el cual cada pregunta tiene* 5 *posibles respuestas, de las cuales únicamente una es la correcta. Si el estudiante conoce la respuesta correcta, esa es la que selecciona, en otro caso selecciona al azar una de las* 5 *posibles respuestas. Supongamos que el estudiante conoce la respuesta correcta de* 70 % *de las preguntas. Dada una pregunta que es contestada correctamente, ¿cuál es la probabilidad de que el estudiante conozca la respuesta correcta?*

Ejercicio 4.22. *Un bolso contiene* 3 *monedas, una de las cuales está acuñada con dos caras mientras que las otras dos monedas son normales (cara de un lado, cruz del otro). Se escoge al azar una moneda del bolso y se lanza* 4 *veces en forma sucesiva. Sabiendo que se obtiene cara en las* 4 *ocasiones que se lanza la moneda, ¿cuál es la probabilidad de que se seleccione la moneda con dos caras?*

Ejercicio 4.23. *En una fábrica hay dos máquinas, A y B, las cuales producen el* 40 % *y el* 60 % *respectivamente de la producción total. Se sabe que la máquina A produce aproximadamente* 3 % *de artículos defectuosos, mientras que la máquina B produce aproximadamente* 5 % *de artículos defectuosos. Encuentre la probabilidad de que un determinado artículo defectuoso sea producido por la máquina A.*

Ejercicio 4.24. *La probabilidad de que cada artículo producido por una máquina satisfaga el estándar es igual a* 0.96*. Cuando un artículo satisface el estándar, la probabilidad de que pase el sistema de inspección es igual a* 0.98*, mientras que cuando no lo satisface la probabilidad es de* 0.05*. Dado que un artículo pasa el sistema de inspección, ¿cuál es la probabilidad de que satisfaga el estándar?*

Ejercicio 4.25. *Supongamos que una prueba de sangre de una cierta enfermedad contagiosa es tal que las probabilidades de un falso negativo y de un falso positivo son ambas iguales a* 0.01*. Supongamos además que la probabilidad de que una persona elegida al azar esté contagiada es igual a p. ¿Para qué valores de p, están contagiadas más del* 90 % *de la personas para las cuales la prueba resulta positiva?*

Ejercicio 4.26. *En un juicio, se está demandando a un hombre pensión alimenticia para un niño, por ser el supuesto padre; sin embargo, el demandado asegura no ser el padre del niño. El abogado de la demandante argumenta que se ha encontrado en el niño una característica genética que posee el padre, la cual se presenta únicamente en el* 1 % *de la población adulta y, cuando está presente, se transmite con seguridad de padre a hijo. El abogado añade que, a priori no se sabe si el hombre demandado es el padre del niño, de manera que la probabilidad de que lo sea es igual a $\frac{1}{2}$, pero,* *dado que se ha encontrado en el niño la característica genética que posee el hombre, la probabilidad de que el demandado sea realmente el padre del niño es casi igual a* 1. *a) Encuentre la probabilidad cercana a* 1 *a la cual hace referencia el abogado. b) ¿Qué se podría argumentar en contra del razonamiento que hace el abogado de la demandante? c) Supongamos que, antes de disponer de la información sobre la presencia en el niño de la característica genética que posee el hombre, el juez determina que la probabilidad de que el hombre sea el padre del niño es igual a* p. *¿Cuál es la nueva probabilidad* P *de que el hombre sea el padre del niño que determinaría el juez una vez que dispone de la nueva información? d) Para* $0 \leq p \leq 0.1$, *grafique* P *como función de* p. *e) Determine los valores de* p *para los cuales* P *resulta cercana a* 1 *(tómese como cercana a* 1 *una probabilidad mayor que* 0.99*).*

EJERCICIO 4.27. *Un médico llegó a la conclusión de que la probabilidad de que su paciente, el Sr. Flores, padezca una cierta enfermedad es igual a* 0.6. *El médico sigue una regla según la cual cuando la probabilidad de que un paciente padezca dicha enfermedad es* 0.8 *o más, entonces recomienda cirugía, pero si la probabilidad es menor a* 0.8, *entonces recomienda estudios adicionales, los cuales son costosos y dolorosos. Por tal motivo, el médico recomienda al Sr. Flores hacerse un estudio, resultando positivo. Dicho estudio tiene la característica de resultar positivo en todos los casos en que la persona padece la enfermedad y negativo en los casos en que la persona no padece la enfermedad y el resultado del estudio no es alterado por otros padecimientos del paciente. El resultado del estudio indicaba al médico que debe de recomendar cirugía, sin embargo, después de realizado el estudio, el Sr. Flores informa al médico que padece de diabetes, lo cual no había contemplado el médico hasta ese momento. Esta información pone en duda al médico pues el estudio que le practicaron al Sr. Flores resulta positivo en* 30% *de los casos en que el paciente no padece la enfermedad pero es diabético. ¿Qué debo recomendar?, se pregunta el médico, ¿otro estudio o cirugía?*

EJERCICIO 4.28. *Un dado desbalanceado está hecho de tal forma que la probabilidad de obtener el número* k *es igual a* ck, *en donde* c *es una constante y* $k \in \{1, 2, 3, 4, 5, 6\}$. *Un experimento aleatorio consiste en lanzar dicho dado; si se obtiene como resultado el número* k, *se lanza una moneda balanceada* k *veces. Dado que al lanzar la moneda se obtiene por lo menos una cara, ¿cuál es la probabilidad de que se obtenga el número* 2 *al lanzar el dado?*

EJERCICIO 4.29. *Un dado balanceado es lanzado* 5 *veces. Después de cada lanzamiento se coloca en una urna* A *una bola blanca cuando el resultado es el número* 6 *y una bola roja cuando el resultado no es el número* 6. *Inmediatamente después, se seleccionan, al azar y con reemplazo,* 3 *bolas de la urna. Dado que las* 3 *bolas seleccionadas son blancas, ¿cuál es la probabilidad de que la urna contenga únicamente bolas blancas?*

EJERCICIO 4.30. *Se tienen dos urnas, la primera contiene* 4 *bolas rojas y* 6 *bolas blancas, mientras que la segunda contiene* 8 *rojas y* 2 *blancas. Se seleccionan, al azar y sin reemplazo,* 2 *bolas de la primera urna y se transfieren a la segunda. Inmediatamente después, se seleccionan, al azar y sin reemplazo,* 2 *bolas de la segunda urna. Sabiendo que finalmente se obtienen* 2 *bolas blancas, ¿cuál es la probabilidad de que se transfieran* 2 *bolas blancas de la primera urna a la segunda?*

EJERCICIO 4.31. *Una moneda balanceada es lanzada* 5 *veces. En cada lanzamiento se coloca en una urna una bola blanca cuando el resultado es cara y una bola roja cuando el resultado es cruz. En seguida se sacan con reemplazo* n *bolas de la urna. Calcule la probabilidad de que la urna contenga únicamente bolas blancas bajo la hipótesis de que las* n *bolas extraídas son blancas.*

EJERCICIO 4.32. *Supongamos que disponemos de* 5 *urnas de tal manera que* 4 *de ellas contienen* 3 *bolas rojas y* 3 *bolas blancas cada una, mientras que la quinta contiene* 5 *bolas blancas y* 1 *bola roja. Un experimento aleatorio consiste en seleccionar, al azar y sin reemplazo,* 3 *bolas de una de las* 5 *urnas, la cual se selecciona también al azar. Sabiendo que las* 3 *bolas seleccionadas son blancas, ¿cuál es la probabilidad de que se seleccione la urna con* 5 *bolas blancas?*

CAPÍTULO 5

LA ADITIVIDAD NUMERABLE

¿No hay entonces ningún fundamento para la oposición que habíamos creído discernir entre el Cálculo de Probabilidades y el individualismo? Por el contrario, hay uno muy real pues el individualismo es antisocial mientras que el Cálculo de Probabilidades es la base de lo que se podría llamar las matemáticas sociales. En efecto, su estudio nos recuerda que vivimos en sociedad y que los fenómenos sociales tienen una existencia real y un interés propio. Nos recuerda que, si bien los hombres son diferentes en muchas cosas, son parecidos en el sentido de que todos ellos están expuestos a los accidentes, a la enfermedad, a la muerte; ... El estudio de esos hechos no puede más que contribuir a desarrollar la noción de solidaridad, a recordar a cada uno que no debe considerarse como independiente del medio en donde vive y que debe participar en la reparación de los daños fortuitos que sufra su vecino y que podría estarlos sufriendo él mismo.

Félix Édouard Justin Émile Borel

5.1. Espacios muestrales infinitos numerables

No es dificil imaginar un experimento aleatorio el cual admita un número infinito de posibles resultados; consideremos por ejemplo el experimento consistente en lanzar una moneda tantas veces como sea necesario hasta obtener una cara; en este caso, dado cualquier número entero positivo n, es posible que la primera cara ocurra en el n-simo lanzamiento. De hecho ya hemos considerado anteriormente experimentos aleatorios de este tipo (véanse los ejemplos 2.34 y 4.6 y los ejercicios 2.32 y 2.33) y hemos calculado probabilidades de eventos asociados con ellos, mediante las reglas y métodos que hemos considerado hasta estos momentos. Sin embargo, si se revisan esos ejemplos se observará que, si bien el experimento aleatorio correspondiente admite una infinidad de posibles resultados, nos interesaba calcular la probabilidad de un evento cuya ocurrencia o no ocurrencia depende solo de un número finito de ellos.

En esta sección consideraremos algunos problemas relacionados con experimentos aleatorios que admiten una infinidad de posibles resultados, pero, a diferencia de los ejemplos tratados con anterioridad, en esta ocasión se tratará de calcular probabilidad de eventos cuya ocurrencia o no ocurrencia depende de una colección infinita de posibles resultados. El tratamiento de estos problemas nos llevará a la consideración de la extensión de la propiedad de la aditividad finita al caso infinito numerable.

Consideremos nuevamente el experimento consistente en lanzar una moneda tantas veces como sea necesario hasta obtener una cara.

Llamando E (éxito) al hecho de obtener cara en un lanzamiento y F (fracaso) al hecho de no obtenerla, cada posible resultado de este experimento puede representarse ya sea mediante una sucesión finita $(F, \ldots, F, S)$ formada por fracasos consecutivos seguidos de un éxito, o bien mediante una sucesión infinita $(F, F, \ldots)$ formada exclusivamente por fracasos. De esta manera, el espacio muestral de este experimento se puede representar de la siguiente manera:

$$\Omega = \{(E), (F, E), (F, F, E), \ldots\} \cup \{(F, F, \ldots)\}$$

Como los lanzamientos son independientes uno del otro, la probabilidad de cada uno de los posibles resultados del experimento se puede calcular fácilmente. Efectivamente, si llamamos ω_n al elemento de Ω consistente de una secuencia de $n-1$ fracasos seguidos de un éxito, se tiene:

$$P(\{\omega_n\}) = \left(\tfrac{1}{2}\right)^n, \text{ para } n \in \mathbb{N}$$

Por otra parte, si denotamos por ω_∞ al elemento $(F, F, F, \ldots)$, entonces la ocurrencia del evento $\{\omega_\infty\}$ implica que, dado cualquier $n \in \mathbb{N}$, en los primeros n lanzamientos se obtiene fracaso, de manera que, aplicando las propiedades de la función de probabilidad, se debe de tener:

$$P(\{\omega_\infty\}) \leq \left(\tfrac{1}{2}\right)^n$$

para toda $n \in \mathbb{N}$. Así que:

$$P(\{\omega_\infty\}) = 0$$

Obsérvese que en el caso de un experimento aleatorio con un espacio muestral finito, $P(A) = 0$ significa que el evento A nunca ocurre; en cambio aquí tenemos un evento de probabilidad cero cuya ocurrencia es posible.

Gracias a la propiedad de la aditividad finita, la probabilidad de cualquier evento compuesto por un número finito de elementos de Ω queda entonces únicamente determinada y se puede calcular simplemente sumando las probabilidades de los posibles resultados que lo componen. De la misma manera, la probabilidad de un evento compuesto por todos los elementos de Ω, excepto un número finito de ellos, también queda únicamente determinada y se puede calcular restando de 1 la suma de las probabilidades de los posibles resultados que no pertenecen al evento en consideración. Sin embargo, el cálculo de la probabilidad de un evento tal que tanto él como su complemento estén compuestos por una infinidad de posibles resultados no puede calcularse inmediatamente aplicando la propiedad de la aditividad finita. Esta probabilidad quedaría únicamente determinada y se podría calcular si agregáramos como propiedad del modelo que queremos construir una análoga a la de la aditividad finita para el caso de una colección infinita numerable de eventos, la cual llamaremos la propiedad de la aditividad numerable.

Para ser más precisos, definamos formalmente el concepto de aditividad numerable y, dado que más adelante utilizaremos este concepto de una manera general, sin hacer referencia a un espacio de probabilidad, conviene dar la definición de este concepto así como del de la aditividad finita dentro de un contexto más amplio.

DEFINICIÓN 5.1 (**Función finitamente aditiva**). *Sea Ω un conjunto y $\mathcal{A}$ la familia de todos los subconjuntos de Ω. Se dice que una función $P : \mathcal{A} \mapsto [0, 1]$ es finitamente aditiva si dada cualquier familia finita de subconjuntos de Ω, $A_1, \ldots, A_n$, tales que $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces $P(\bigcup\limits_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$.*

DEFINICIÓN 5.2 (**Función $\boldsymbol{\sigma}$-aditiva**). *Sea Ω un conjunto y $\mathcal{A}$ la familia de todos los subconjuntos de Ω. Se dice que una función $P : \mathcal{A} \mapsto [0, 1]$ es σ-aditiva, o que satisface la propiedad de la aditividad numerable, si es finitamente aditiva y dada cualquier familia infinita numerable de subconjuntos de Ω, $A_1, A_2, \ldots$, tales que $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces $P(\bigcup\limits_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$.*

La propiedad de la aditividad numerable va a jugar el mismo papel que las otras propiedades de la función de probabilidad, es decir, servirá para poder extender la función de probabilidad a una familia más grande de eventos cuando se ha iniciado el proceso de asignar probabilidades a una familia particular. Con algunas consideraciones adicionales sobre la familia de eventos a los cuales se podrá asignar una probabilidad, en un modelo matemático general, **la propiedad de la aditividad numerable**

deberá agregarse como propiedad adicional de la función de probabilidad si queremos que ésta se pueda extender a eventos que dependen de una colección infinita de posibles resultados. Sin embargo, es ilustrativo abundar en el estudio de esta propiedad pues se puede mostrar que en algunos casos se puede obtener como consecuencia de las otras propiedades de la función de probabilidad que hemos estudiado hasta este momento.

Ejemplo 5.3. *Supongamos que, dado el experimento aleatorio descrito antes, queremos calcular la probabilidad de que el experimento se termine en un número par de lanzamientos. El correspondiente evento A se puede entonces representar como $A = \{\omega_2, \omega_4, \ldots\}$. De manera que, asumiendo que la función de probabilidad es σ-aditiva, se tiene:*

$P(A) = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{2k} = \frac{1}{3}$

La probabilidad del evento A se puede encontrar con otro método, aparentemente sin apelar a una generalización de la propiedad de la aditividad finita al caso infinito. En efecto, consideremos los siguientes eventos:

A_1: se obtiene fracaso en el primer lanzamiento y éxito en el segundo.

C: se obtiene fracaso en los primeros dos lanzamientos.

Por la propiedad de la aditividad finita, se tiene:

$P(A) = P(A_1) + P(A \cap C) = P(A_1) + P(A \mid C)P(C) = P(A_1) + P(A)P(C)$

Pero, dado que C ocurre, las condiciones del experimento son como las del inicio, por lo tanto, $P(A \mid C) = P(A)$. Así que:

$P(A) = P(A_1) + P(A \mid C)P(C) = P(A_1) + P(A)P(C)$

Por lo tanto:

$P(A) = \frac{P(A_1)}{1-P(C)} = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{1}{3}$

Aparentemente no se está haciendo aquí referencia a la propiedad de σ-aditividad; sin embargo, vamos a mostrar que la relación $P(A \cap C) = P(A)P(C)$ la contiene implícitamente. En efecto, definamos, para $k \in \mathbb{N}$, $Q_k = \{\omega_k, \omega_{k+2}, \ldots\}$. La relación $P(A \cap C) = P(A)P(C)$ se puede escribir entonces como $P(Q_4) = P(C)P(Q_2)$.

Pero, para $k \in \{3, 4, \ldots\}$, tenemos $P(\{\omega_k\}) = \frac{1}{4}P(\{\omega_{k-2}\})$. Así que, utilizando la propiedad de la aditividad finita, para $k \in \{2, 3, \ldots\}$, se tiene:

$P(Q_{2k}) = \frac{1}{4}P(Q_{2(k-1)})$

Por lo tanto, $P(Q_{2k}) = \left(\frac{1}{4}\right)^{k-1} P(Q_2)$ *y entonces* $\lim_{k \rightsquigarrow \infty} P(Q_{2k}) = 0$.

Ahora, para $k \in \mathbb{N}$, *definamos* $A_k = \{\omega_2, \omega_4, \ldots, \omega_{2k}\}$. *Entonces, utilizando la propiedad de la aditividad finita, se tiene:*

$$P(A) = P(A_k) + P(Q_{2(k+1)}) = \sum_{\omega \in A_k} p(\omega) + P(Q_{2(k+1)})$$

Así que, tomando límites, se tiene:

$$P(A) = \lim_{k \rightsquigarrow \infty} \sum_{\omega \in A_k} p(\omega) = \sum_{\omega \in A} p(\omega)$$

Se concluye entonces que el razonamiento basado en probabilidades condicionales contiene implícitamente la propiedad de σ*-aditividad.*

Todo indica entonces que un problema del tipo planteado únicamente se puede resolver agregando a las propiedades de la función de probabilidad la de σ*-aditividad. Sin embargo, vamos a ver que, en este caso particular, esto no es necesario pues la probabilidad del evento* A *puede obtenerse recurriendo únicamente a la aditividad finita y a las otras propiedades básicas de la función de probabilidad.*

Sea B *el evento 'el experimento se termina en un número impar de lanzamientos', entonces* $B = \{\omega_1, \omega_3, \ldots\}$. *Definamos además* $p_n = P(\{\omega_n\})$.

Para cualquier $n \in \mathbb{N}$, *se tiene* $\{\omega_2, \omega_4, \ldots, \omega_{2n}\} \subset A$ *y* $\{\omega_1, \omega_3, \ldots, \omega_{2n-1}\} \subset B$, *por lo tanto* $P(A) \geq p_2 + p_4 + \cdots + p_{2n}$ *y* $P(B) \geq p_1 + p_3 + \cdots + p_{2n-1}$. *De manera que, tomando límites, cuando* $n \rightsquigarrow \infty$, *se tiene:*

$$P(A) \geq \sum_{n=1}^{\infty} p_{2n} = \sum_{n=1}^{\infty} \frac{1}{2^{2n}} = \frac{1}{3}$$

$$P(B) \geq \sum_{n=1}^{\infty} p_{2n-1} = \sum_{n=1}^{\infty} \frac{1}{2^{2n-1}} = \frac{2}{3}$$

Supongamos que $P(A) > \frac{1}{3}$, *entonces se tendría:*

$$1 \geq P(A) + P(B) > \frac{1}{3} + \frac{2}{3} = 1$$

lo cual es una contradicción, por lo tanto, $P(A) = \frac{1}{3}$ *y* $P(B) = \frac{2}{3}$.

▲

Vamos a ver en seguida que en realidad para un problema como el planteado en el ejemplo anterior, la propiedad de σ-aditividad es una consecuencia de las 3 propiedades básicas de la función de probabilidad que hemos considerado hasta ahora. Este resultado no debe interpretarse en el sentido de que siempre la σ-aditividad se puede obtener como consecuencia de estas propiedades. El resultado se probará únicamente para una clase particular de experimentos aleatorios. Se puede mostrar que, **en general, la σ-aditividad no es consecuencia de las 3 propiedades básicas de la función de probabilidad consideradas hasta ahora y entonces, si queremos ampliar la familia de eventos a los cuales se les puede asignar una probabilidad de manera única, la propiedad de σ-aditividad se debe de agregar como una propiedad adicional**.

TEOREMA 5.4. *Sea $\Omega = \{\omega_1, \omega_2, \ldots\}$ un conjunto infinito numerable, $\mathcal{A}$ la familia formada por todos los subconjuntos de Ω y $P : \mathcal{A} \mapsto [0,1]$ una función que satisface las siguientes condiciones:*

(i) *P es finitamente aditiva*
(ii) $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$

entonces P es σ-aditiva.

Demostración

Definamos $p_k = P(\{\omega_k\})$ y sea A cualquier subconjunto de Ω infinito numerable, digamos $A = \{\omega_{j_1}, \omega_{j_2}, \ldots\}$. Sea $A^c = \{\omega_{i_1}, \omega_{i_2}, \ldots\}$. En vista de que P es no negativa y finitamente aditiva, es monótona no decreciente, así que $P(A) \geq p_{j_1} + \ldots + p_{j_n}$ para toda $n \in \mathbb{N}$. Por lo tanto, $P(A) \geq \sum_{k=1}^{\infty} p_{j_k}$. De la misma manera, se obtiene $P(A^c) \geq \sum_k p_{i_k}$. Además, como la serie $\sum_{k=1}^{\infty} p_k$ converge a 1, las dos series $\sum_{k=1}^{\infty} p_{j_k}$ y $\sum_k p_{i_k}$ son también convergentes y su suma es 1. Ahora bien, si $P(A) > \sum_{k=1}^{\infty} p_{j_k}$, entonces tendríamos:

$1 \geq P(\Omega) = P(A) + P(A^c) > \sum_{k=1}^{\infty} p_{j_k} + \sum_k p_{i_k} = 1$

lo cual es una contradicción. Por lo tanto, podemos concluir que $P(A) = \sum_{k=1}^{\infty} p_{j_k}$.

Sea ahora $A_1, A_2, \ldots$ una colección infinita numerable de subconjuntos no vacíos de Ω tales que $A_i \cap A_j = \emptyset$ para $i \neq j$. Si $A = \cup_{n=1}^{\infty} A_n = \{\omega_{k_1}, \omega_{k_2}, \ldots\}$, entonces, como P es no negativa y finitamente aditiva, se tiene $P(A) \geq \sum_{n=1}^{n_0} P(A_n)$ para toda $n_0 \in \mathbb{N}$. Así que:

$$P(A) \geq \sum_{n=1}^{\infty} P(A_n) \tag{5.1}$$

Por otro lado, dada $N \in \mathbb{N}$, existe $n_0 \in \mathbb{N}$ tal que $\{\omega_{k_j} : j \leq N\} \subset \cup_{n=1}^{n_0} A_n$, así que:

$\sum_{j=1}^{N} p_{k_j} = P(\{\omega_{k_j} : j \leq N\}) \leq P(\bigcup_{n=1}^{n_0} A_n) = \sum_{n=1}^{n_0} P(A_n) \leq \sum_{n=1}^{\infty} P(A_n)$

Es decir, $\sum_{j=1}^{N} p_{k_j} \leq \sum_{n=1}^{\infty} P(A_n)$ para toda $N \in \mathbb{N}$. Por lo tanto:

$$P(A) = \sum_{j=1}^{\infty} p_{k_j} \leq \sum_{n=1}^{\infty} P(A_n) \tag{5.2}$$

Combinando (5.1) y (5.2), obtenemos $P(A) = \sum_{n=1}^{\infty} P(A_n)$, lo cual prueba el resultado.

■

COROLARIO 5.5. *Sea Ω cualquier conjunto, $\mathcal{A}$ la familia formada por todos los subconjuntos de Ω y $P : \mathcal{A} \mapsto [0,1]$ una función finitamente aditiva. Supongamos que existe un subconjunto numerable Ω' de Ω tal que $\sum_{\omega \in \Omega'} P(\{\omega\}) = 1$. Entonces P es σ-aditiva y $P(A) = \sum_{\{\omega \in A \cap \Omega'\}} P(\{\omega\})$ para todo $A \in \mathcal{A}$.*

Demostración

Sea $\mathcal{A}'$ la familia formada por todos los subconjuntos de Ω'. Entonces, de acuerdo con el teorema 5.4, restringida a $\mathcal{A}'$, P es σ-aditiva. En particular:

$P(\Omega') = \sum_{\omega \in \Omega'} P(\{\omega\}) = 1.$

Así que, si B es cualquier subconjunto de $\Omega - \Omega'$, tenemos:

$1 \geq P(B \cup \Omega') = P(B) + P(\Omega') = P(B) + 1$

de lo cual se sigue $P(B) = 0$. Entonces, si A es cualquier subconjunto de Ω, se tiene $P(A) = P(A \cap \Omega')$. Así que, si $A_1, A_2, \ldots$ es cualquier colección numerable de subconjuntos de Ω tales que $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces:

$P(\bigcup_k A_k) = P[(\bigcup_k A_k) \cap \Omega'] = P[\bigcup_k (A_k \cap \Omega')] = \sum_k P(A_k \cap \Omega') = \sum_k P(A_k)$

lo cual prueba que P es σ-aditiva sobre $\mathcal{A}$.

Finalmente, obsérvese que, en particular, si A es cualquier subconjunto de Ω, se tiene:

$P(A) = P(A \cap \Omega') = \sum_{\{\omega \in A \cap \Omega'\}} P(\{\omega\})$

■

Vamos ahora a aplicar este resultado a una clase de experimentos aleatorios, la cual incluye el ejemplo con el que se inició esta sección y otros similares.

DEFINICIÓN 5.6 (**Ensayos de Bernoulli**). *Un ensayo de Bernoulli es un experimento aleatorio el cual admite únicamente dos posibles resultados, los cuales llamaremos éxito y fracaso, respectivamente.*

Sea $B_1, B_2, \ldots$ una sucesión de ensayos de Bernoulli independientes y $r_k > 0$ la probabilidad de éxito en el k-ésimo ensayo.

Sea Φ el conjunto formado por todas las colecciones finitas $(s_1, \ldots, s_n)$ de 0's y 1's y, para cada $\omega = (s_1, \ldots, s_n) \in \Phi$, definamos $p(\omega) = p(s_1) \cdots p(s_n)$, en donde:

$$p(s_k) = \begin{cases} r_k & \text{si} \quad s_k = 1 \\ 1 - r_k & \text{si} \quad s_k = 0 \end{cases}$$

Si $\omega = (s_1, \ldots, s_n) \in \Phi$, diremos que ω tiene longitud n y definimos $l(\omega) = n$.

Consideremos el experimento aleatorio $\mathcal{E}$ consistente en la realización consecutiva de los ensayos de Bernoulli $B_1, B_2, \ldots$ y supongamos que está definida una condición $\mathcal{C}$ de tal manera que el experimento se termina cuando la condición $\mathcal{C}$ se satisface por primera vez. Por ejemplo, $\mathcal{C}$ podría definirse como la ocurrencia de cierto número fijo de éxitos.

El espacio muestral Ω de $\mathcal{E}$ puede descomponerse en dos subconjuntos Ω^F y Ω^I, en donde Ω^F es el conjunto de resultados de las sucesiones finitas de ensayos de Bernoulli que terminan con la primera ocurrencia de $\mathcal{C}$ y Ω^I es el conjunto de resultados de las sucesiones infinitas de ensayos de Bernoulli para las cuales la condición $\mathcal{C}$ nunca se alcanza en un número finito de realizaciones de los ensayos de Bernoulli. Cada elemento de Ω^F puede representarse mediante una sucesión finita $(s_1, \ldots, s_n)$ de 0's y 1's y cada elemento de Ω^I puede representarse mediante una sucesión infinita $(s_1, s_2, \ldots)$ de 0's y 1's, en donde $s_i = 0$ y $s_i = 1$ representan, respectivamente, un fracaso y un éxito en el i-ésimo ensayo. En vista de que Φ es un conjunto numerable y $\Omega^F \subset \Phi$, Ω^F es también un conjunto numerable. Sin embargo, Ω^I puede ser un conjunto no numerable.

Lema 5.7. *$\sum_{\{\omega \in \Phi : l(\omega) = n\}} p(\omega) = 1$ para cualquier $n \in \mathbb{N}$.*

Demostración

Para cada $n \in \mathbb{N}$, definamos $S(n) = \sum_{\{\omega \in \Phi : l(\omega) = n\}} p(\omega)$. Entonces , $S(1) = r_1 + (1 - r_1) = 1$ y $S(n+1) = r_{n+1} S(n) + (1 - r_{n+1}) S(n) = S(n)$ para toda $n \in \mathbb{N}$.

■

Lema 5.8. *La serie $\sum_{\omega \in \Omega^F} p(\omega)$ siempre es convergente y $\sum_{\omega \in \Omega^F} p(\omega) \leq 1$.*

Demostración

Para cada $n \in \mathbb{N}$, definamos:

$$\Omega^F_{(n)} = \left\{ \omega \in \Omega^F \mid l(\omega) \leq n \right\}$$

y, para cada $\omega \in \Omega^F_{(n)}$, definamos:

$\Phi^{(n)}(\omega) = \{\omega' \in \Phi \mid l(\omega') = n \text{ y } \omega' \text{ comienza con } \omega\}$

Dado $\omega \in \Omega^F_{(n)}$, sea $m = l(\omega)$. Si $m = n$, entonces $\Phi^{(n)}(\omega) = \{\omega\}$, si no:

$\sum_{\omega' \in \Phi^{(n)}(\omega)} p(\omega') = p(\omega) \sum_{\{\omega' \in \Phi : l(\omega') = n-m\}} p(\omega') = p(\omega)$

Así que, en cualquier caso:

$p(\omega) = \sum_{\omega' \in \Phi^{(n)}(\omega)} p(\omega')$

Obviamente, dados dos elementos distintos $\omega, \omega' \in \Omega^F_{(n)}$, los subconjuntos $\Phi^{(n)}(\omega)$ y $\Phi^{(n)}(\omega')$ son ajenos. Además:

$\bigcup_{\omega \in \Omega^F_{(n)}} \Phi^{(n)}(\omega) \subset \{\omega \in \Phi \mid l(\omega) = n\}$

Así que, para cualquier $n \in \mathbb{N}$, tenemos:

$\sum_{\omega \in \Omega^F_{(n)}} p(\omega) \leq \sum_{\{\omega \in \Phi : l(\omega) = n\}} p(\omega) = 1$

Por lo tanto, la serie:

$\sum_{\omega \in \Omega^F} p(\omega) = \text{lím}_{n \rightsquigarrow \infty} \sum_{\omega \in \Omega^F_{(n)}} p(\omega)$

es convergente y $\sum_{\omega \in \Omega^F} p(\omega) \leq 1$.

■

Proposición 5.9. *Dada $M \in \mathbb{N}$ y $t_1, \ldots, t_M$ M números naturales fijos, supongamos que los ensayos de Bernoulli se clasifican en M tipos, $T_1, \ldots, T_M$, con probabilidades de éxito $p_1, \ldots, p_M$, respectivamente, y que la condición $\mathcal{C}$ se satisface cuando, para alguna $i \in \{1, \ldots, M\}$, se obtienen t_i éxitos en los ensayos de Bernoulli de tipo T_i. Sea Ω el espacio muestral del correspondiente experimento aleatorio, Ω^F el conjunto de resultados de las sucesiones finitas de ensayos de Bernoulli que terminan con la primera ocurrencia de $\mathcal{C}$, $\mathcal{A}$ la familia de todos los subconjuntos de Ω y $P : \mathcal{A} \mapsto [0,1]$ una función de probabilidad (finitamente aditiva) tal que $P(\{\omega\}) = p(\omega)$ para toda $\omega \in \Omega^F$. Entonces, P es σ-aditiva y, además, $P(A) = \sum_{\{\omega \in A \cap \Omega^F\}} p(\omega)$ para todo $A \in \mathcal{A}$.*

Demostración

De acuerdo con el corolario 5.5, basta con demostrar que $\sum_{\{\omega \in \Omega^F\}} p(\omega) = 1$.

Sea $t = t_1 + \cdots + t_M$, $p = \text{mín}\{p_1, \ldots, p_M\}$ y, para cada $n \in \mathbb{N}$, definamos el evento:

A_{nt}: La primera ocurrencia de $\mathcal{C}$ se obtiene en alguno de los primeros nt ensayos.

A_{nt} se representa mediante un conjunto finito formado por los elementos de Ω^F de longitud menor o igual a nt, de manera que:

$P(A_{nt}) = \sum_{\{\omega\in\Omega^F:l(\omega)\leq nt\}} p(\omega)$

Por lo tanto:

$\sum_{\{\omega\in\Omega^F\}} p(\omega) = \lim_{n\rightsquigarrow\infty} \sum_{\{\omega\in\Omega^F:l(\omega)\leq nt\}} p(\omega) = \lim_{n\rightsquigarrow\infty} P(A_{nt})$

Vamos a demostrar que $P\left(A_{nt}^c\right) \leq (1-p^t)^n$ para cualquier $n \in \mathbb{N}$, de lo cual se sigue que $\lim_{n\rightsquigarrow\infty} P(A_{nt}) = 1$, así que la proposición quedará demostrada.

La demostración de la desigualdad $P\left(A_{nt}^c\right) \leq (1-p^t)^n$, para cualquier $n \in \mathbb{N}$, se hará por inducción sobre n.

Si en los primeros t ensayos hay éxito, entonces ocurre A_t, así que $P\left(A_t\right) \geq p^t$. Por lo tanto, $P\left(A_t^c\right) \leq 1-p^t$.

Obsérvese que si $\alpha = (s_1,\ldots,s_m) \in A_{nt}$, entonces, por el lema 5.7, se tiene:

$p(\alpha) = \sum_{\{\omega\in\Phi:l(\omega)=nt,\ \omega \text{ y } \alpha \text{ coinciden en sus primeros } m \text{ elementos}\}} p(\omega)$

Así que, si:

$$C_{nt} = \left\{\omega \in \Phi : \begin{array}{l} l(\omega) = nt \text{ y los primeros elementos de } \omega \text{ no} \\ \text{coinciden con los elementos de ninguna } \omega' \in A_{nt} \end{array}\right\}$$

entonces $P(A_{nt}^c) = \sum_{\{\omega\in C_{nt}\}} p(\omega)$.

Supongamos ahora que la desigualdad $P\left(A_{nt}^c\right) \leq (1-p^t)^n$ se cumple para $n = k \in \mathbb{N}$, entonces:

$P\left(A_{(k+1)t}^c\right) = P\left(A_{(k+1)t}^c \cap A_{kt}^c\right) = P\left(A_{kt}^c\right) - P\left(A_{(k+1)t} \cap A_{kt}^c\right)$

$A_{(k+1)t} \cap A_{kt}^c$ está formada por los elementos de Ω^F de longitud mayor que kt y menor o igual a $(k+1)t$, de manera que:

$P(A_{(k+1)t} \cap A_{kt}^c) = \sum_{\{\omega\in\Omega^F:kt<l(\omega)\leq(k+1)t\}} p(\omega)$

Pero, si $\alpha = (s_1,\ldots,s_m) \in A_{(k+1)t} \cap A_{kt}^c$, entonces, por el lema 5.7, se tiene:

$p(\alpha) = \sum_{\{\omega\in\Phi:l(\omega)=(k+1)t,\ \omega \text{ y } \alpha \text{ coinciden en sus primeros } m \text{ elementos}\}} p(\omega)$

Así que, si:

$$D_{kt} = \left\{ \omega \in \Phi : \begin{array}{l} l(\omega) = (k+1)t, \text{ los primeros } kt \text{ elementos de } \omega \text{ coinciden con} \\ \text{los elementos de alguna } \omega' \in C_{kt} \text{ y los primeros elementos de} \\ \omega \text{ coinciden con los elementos de alguna } \omega' \in A_{(k+1)t} \end{array} \right\}$$

$$E_k t = \left\{ \omega \in \Phi : \begin{array}{l} l(\omega) = (k+1)t \text{ y } \omega \text{ está formada por} \\ \text{una sucesión en } C_{kt} \text{ seguida de } t \text{ 1's} \end{array} \right\}$$

entonces:

$$P(A_{(k+1)t} \cap A_{kt}^c) = \sum_{\{\omega \in D_{kt}\}} p(\omega) \geq \sum_{\{\omega \in E_{kt}\}} p(\omega) \geq p^t \sum_{\{\omega \in C_{kt}\}} p(\omega) = p^t P(A_{kt}^c)$$

Así que:

$$P\left(A_{(k+1)t}^c\right) = P\left(A_{kt}^c\right) - P\left(A_{(k+1)t} \cap A_{kt}^c\right) \leq P\left(A_{kt}^c\right) - p^t P\left(A_{kt}^c\right) = (1-p^t)P\left(A_{kt}^c\right) \leq (1-p^t)^{k+1}$$

∎

Comentario 5.10. *La última proposición muestra que si queremos construir un modelo probabilístico para el experimento aleatorio $\mathcal{E}$ de tal manera que se satisfagan las siguientes condiciones:*

(i) *Cualquier subconjunto del espacio muestral representa un evento,*
(ii) *La función de probabilidad $P : \mathcal{A} \mapsto [0,1]$, en donde $\mathcal{A}$ es la familia de todos los subconjuntos de Ω, es finitamente aditiva,*
(iii) *$P(\{\omega\}) = p(\omega)$ para toda $\omega \in \Omega^F$.*

Entonces, ***únicamente hay una posible elección para la función de probabilidad P****, a saber, $P(A) = \sum_{\omega \in A \cap \Omega^F} P(\{\omega\})$ para cualquier $A \in \mathcal{A}$, lo cual define una función de probabilidad σ-aditiva.*

A continuación se presentan algunos problemas que caen dentro de la categoría de problemas descritos en la proposición 5.9 y, por consiguiente, en cada uno de ellos, la asignación de probabilidades dada en el comentario 5.10 define la única posible función de probabilidad satisfaciendo las condiciones *i ii* y *iii* ahí enunciadas.

Ejemplo 5.11. *Una persona* A *juega contra otra persona B lanzando alternadamente un par de dados. A lanza primero los dados, con la condición de que gana el juego si obtiene una suma igual a* 6*; en caso contrario, el juego continúa lanzando B los dados, con la condición de que gana el juego si obtiene una suma igual a* 7*; en* *caso contrario, el juego continúa lanzando A los dados bajo las condiciones iniciales. Calcule las probabilidades que cada jugador tiene de ganar el juego.*[1]

Solución

Vamos a resolver este problema por dos métodos, en el primero se utilizará explícitamente la propiedad de la aditividad numerable[2]*, mientras que en el segundo se utilizará un razonamiento basado en probabilidades condicionales, el cual, al igual que en el caso del ejemplo tratado con anterioridad, contiene implícitamente la propiedad de la aditividad numerable*[3]

1er. método:

Para cada lanzamiento, definamos como éxito al hecho de que el jugador que lanza los dados gana el juego en ese lanzamiento. Entonces cada posible resultado de este juego puede representarse ya sea mediante una sucesión finita $(F, \ldots, F, S)$ *formada por fracasos consecutivos seguidos de un éxito, o bien mediante una sucesión infinita* $(F, F, \ldots)$ *formada exclusivamente por fracasos. Denotemos por* ω_n *al posible resultado* $(F, \ldots, F, S)$ *formado por* $n-1$ *fracasos seguidos de un éxito. Entonces, de acuerdo con las reglas estudiadas en las secciones anteriores, se puede asignar una probabilidad* $p(\omega_n)$ *a cada* ω_n *de la siguiente manera:*

$p(\omega_{2k}) = (1-p_1)^k(1-p_2)^{k-1}p_2$ *si* $n = 2k$ *para alguna* $k \in \mathbb{N}$

$p(\omega_{2k-1}) = (1-p_1)^{k-1}(1-p_2)^{k-1}p_1$ *si* $n = 2k-1$ *para alguna* $k \in \mathbb{N}$

en donde p_1, p_2 *son las probabilidades, respectivamente, de obtener* 6 *y* 7 *puntos al lanzar el par de dados, es decir* $p_1 = \frac{5}{36}$ *y* $p_2 = \frac{1}{6}$. *Ahora bien, como la función de probabilidad es* σ*-aditiva, las probabilidades* $P(A)$ *y* $P(B)$ *de que* A *y* B *ganen, respectivamente, están dadas por:*

$P(A) = \sum_{k=1}^{\infty} p(\omega_{2k-1}) = \sum_{k=1}^{\infty} \left(\frac{31}{36}\right)^{k-1} \left(\frac{5}{6}\right)^{k-1} \frac{5}{36} = \frac{30}{61}$

$P(B) = \sum_{k=1}^{\infty} p(\omega_{2k}) = \sum_{k=1}^{\infty} \left(\frac{31}{36}\right)^{k} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} = \frac{31}{61}$

---

[1]Este problema fue propuesto por Christiaan Huygens a M. de Roberval en el año 1656 y constituye el primer planteamiento de un problema de probabilidad en el cual el conjunto de posibles resultados es infinito ([**7**]).

[2]Este método de solución fue utilizado por Jacques Bernoulli en el año 1705 para resolver este problema ([**3**]).

[3]Este fue básicamente el método utilizado por C. Huygens en su libro "Du Calcul dans les Jeux de Hasard" , publicado en 1657 ([**7**]).

2o. método:

Consideremos los siguientes eventos:

A*: A gana el juego.*

B*: B gana el juego.*

A_1*: Se obtiene éxito en el primer lanzamiento.*

B_1*: Se obtiene fracaso en el primer lanzamiento y éxito en el segundo.*

C*: Se obtiene fracaso en los primeros dos lanzamientos.*

Utilizando la propiedad de la aditividad finita y observando que la ocurrencia de C coloca a los dos jugadores en la misma situación que al inicio del juego, se tiene:

$P(A) = P(A_1) + P(A \cap C) = P(A_1) + P(A \mid C)P(C) = P(A_1) + P(A)P(C)$

$P(B) = P(B_1) + P(B \cap C) = P(B_1) + P(B \mid C)P(C) = P(B_1) + P(B)P(C)$

De lo cual se obtiene, $P(A) = \frac{P(A_1)}{1-P(C)}$ *y* $P(B) = \frac{P(B_1)}{1-P(C)}$. *Esto resuelve el problema ya que* $P(A_1) = \frac{5}{36}$, $P(B_1) = \frac{31}{36}\frac{1}{6}$, *y* $P(C) = \frac{31}{36}\frac{5}{6}$. *Así que ,* $P(A) = \frac{\frac{5}{36}}{1-\frac{31}{36}\frac{5}{6}} = \frac{30}{61}$ *y* $P(B) = \frac{\frac{31}{36}\frac{1}{6}}{1-\frac{31}{36}\frac{5}{6}} = \frac{31}{61}$.

Ejemplo 5.12. *Tres personas, P, Q y R, colocan doce fichas en una urna,* 4 *de ellas son blancas y* 8 *negras. Juegan sacando por turnos una ficha al azar de la caja, reemplazándola después y con la condición de que el primero que saque una ficha blanca gana el juego. Suponiendo que el primer turno es de P, el segundo de Q y el tercero de R, calcule las probabilidades que cada jugador tiene de ganar.*

Solución

Para cada elección, llamemos éxito al hecho de que el jugador que está haciendo la elección obtiene una bola blanca. Entonces cada posible resultado de este experimento puede representarse ya sea mediante una sucesión finita $(F, , F, S)$*, formada por fracasos consecutivos seguidos de un éxito, o bien mediante una sucesión infinita* $(F, F, \ldots)$*, formada exclusivamente por fracasos. Denotemos por* ω_n *al posible resultado* $(F, \ldots, F, S)$ *formado por* $n-1$ *fracasos seguidos de un éxito. Entonces, de acuerdo con las reglas estudiadas en las secciones anteriores, se puede asignar una probabilidad* $p(\omega_n)$ *a cada* ω_n *de la siguiente manera:*

$p(\omega_n) = p(1-p)^{n-1}$

en donde p *es la probabilidad de obtener una bola blanca en una elección, es decir* $p = \frac{1}{3}$*. Ahora bien, como la función de probabilidad es* σ*-aditiva, las probabilidades* $P(A)$*,* $P(B)$ *y* $P(C)$ *de que* P*,* Q *y* R *ganen, respectivamente, están dadas por:*

$$P(A) = \sum_{k=0}^{\infty} p(\omega_{3k+1}) = \sum_{k=0}^{\infty} \frac{1}{3}\left(\frac{2}{3}\right)^{3k} = \frac{9}{19}$$

$$P(B) = \sum_{k=0}^{\infty} p(\omega_{3k+2}) = \sum_{k=0}^{\infty} \frac{1}{3}\left(\frac{2}{3}\right)^{3k+1} = \frac{6}{19}$$

$$P(C) = \sum_{k=0}^{\infty} p(\omega_{3k+3}) = \sum_{k=0}^{\infty} \frac{1}{3}\left(\frac{2}{3}\right)^{3k+2} = \frac{4}{19}$$

Ejemplo 5.13 (**Problema de los 3 jugadores**). *Tres jugadores,* P*,* Q *y* R*, juegan partidas por parejas en cada una de las cuales la probabilidad que cada jugador tiene de ganar es* $\frac{1}{2}$*; quien gane una partida juega con el otro jugador hasta que uno de los jugadores gane dos partidas consecutivas, ganando entonces el juego. Suponiendo que comienzan jugando* P *contra* Q*, calcule las probabilidades que cada uno tiene de ganar el juego.*

Solución

El experimento aleatorio $\mathcal{E}$ *asociado con este problema puede verse como la realización de una sucesión de ensayos de Bernoulli* $B_1, B_2, \ldots$*, en donde* B_1 *es éxito si el jugador* P *gana la primera partida y, para* $i \geq 1$*,* B_{i+1} *es éxito si el jugador que ganó la* i*-ésima partida gana también la que sigue. Con estas convenciones, el juego se termina en el momento en que por primera vez se obtiene éxito después del primer ensayo. Entonces cada posible resultado de este experimento puede representarse ya sea mediante una sucesión finita* $(F, \ldots, F, S)$*, formada por fracasos consecutivos seguidos de un éxito , o bien mediante una sucesión finita* $(S, F, \ldots, F, S)$*, formada por un éxito seguido de fracasos consecutivos a los cuales sigue un éxito, o bien mediante una sucesión infinita* $(F, F, \ldots)$*, formada exclusivamente por fracasos. Denotemos por* ω_n *al posible resultado* $(F, \ldots, F, S)$ *formado por* n *fracasos seguidos de un éxito y por* ω_n' *al posible resultado* $(S, F, \ldots, F, S)$ *formado por un éxito seguido de* $n-1$ *fracasos consecutivos a los cuales sigue un éxito. Entonces, de acuerdo con las reglas estudiadas en las secciones anteriores, se puede asignar probabilidades* $p(\omega_n)$ *y* $p(\varpi_n)$ *a cada* ω_n *y* ω_n' *de la siguiente manera:*

$$p(\omega_n) = p(1-p)^n$$

$$p(\omega_n') = p^2(1-p)^{n-1}$$

en donde p *es la probabilidad de éxito en cada ensayo, es decir* $p = \frac{1}{2}$*. Ahora bien, como la función de probabilidad es* σ*-aditiva, las probabilidades* $P(A)$*,* $P(B)$ *y* $P(C)$ *de que* P*,* Q *y* R *ganen, respectivamente, están dadas por:*

$$P(A) = \sum_{k=0}^{\infty} p(\omega_{3k+3}) + \sum_{k=0}^{\infty} p(\omega_{3k+1}) = \sum_{k=0}^{\infty} \left(\tfrac{1}{2}\right)^{3k+4} + \sum_{k=0}^{\infty} \left(\tfrac{1}{2}\right)^{3k+2} = \tfrac{5}{14}$$

$$P(B) = \sum_{k=0}^{\infty} p(\omega_{3k+1}) + \sum_{k=0}^{\infty} p(\omega_{3k+3}) = \sum_{k=0}^{\infty} \left(\tfrac{1}{2}\right)^{3k+2} + \sum_{k=0}^{\infty} \left(\tfrac{1}{2}\right)^{3k+4} = \tfrac{5}{14}$$

$$P(C) = \sum_{k=0}^{\infty} p(\omega_{3k+2}) + \sum_{k=0}^{\infty} p(\omega_{3k+2}) = \sum_{k=0}^{\infty} \left(\tfrac{1}{2}\right)^{3k+3} + \sum_{k=0}^{\infty} \left(\tfrac{1}{2}\right)^{3k+3} = \tfrac{2}{7}$$

▲

El siguiente problema es especialmente importante por el papel que jugó en el desarrollo de la Teoría de la Probabilidad.

EJEMPLO 5.14 (**Problema de la ruina del jugador**). *Dos jugadores, A y B, los cuales poseen, respectivamente, a y b fichas, juegan partidas consecutivas, en cada una de las cuales las probabilidades de que A y B ganen son, respectivamente, p y $q = 1 - p$ y de tal manera que en cada una de ellas el perdedor dará una de sus fichas al vencedor. Si el ganador del juego es aquel que en algún momento llegue a poseer la totalidad de fichas, ¿qué probabilidad tiene cada jugador de ganar el juego?*[4]

Solución

Obsérvese que este problema no cae dentro de la categoría de problemas descritos en la proposición 5.9, de manera que no podemos utilizarla para probar que, al igual en los otros problemas que se han tratado en esta sección, la propiedad de la aditividad numerable es consecuencia de las otras propiedades de la función de probabilidad. Sin embargo, utilizando el mismo método que el utilizado en la demostración de la proposición 5.9, se puede mostrar que también en este caso la propiedad de la aditividad numerable es consecuencia de las otras propiedades de la función de probabilidad.

El uso directo de la propiedad de la aditividad numerable en la solución de este problema da lugar a una serie cuyo cálculo no es evidente, así que resulta más sencillo aplicar el método basado en probabilidades condicionales, el cual contiene, implícitamente, la propiedad de la aditividad numerable.

Para $z \in \{0, \ldots, a+b\}$, denotemos por p_z a la probabilidad de que el jugador P quede arruinado al comenzar con z fichas. Por la regla de la probabilidad total, tenemos, para $z \in \{1, \ldots, a+b-1\}$:

$$p_z = pp_{z+1} + qp_{z-1}$$

Además, $p_0 = 1$ y $p_{a+b} = 0$.

Esto establece una ecuación en diferencias. Para resolverla la escribiremos en la siguiente forma equivalente, la cual se obtiene escribiendo $p_z = p_z(p+q)$:

[4] Este problema fue propuesto a Pierre de Fermat por Blaise Pascal en el año 1656.

$q\,(p_z - p_{z-1}) = p\,(p_{z+1} - p_z)$

Consideremos dos casos. Supongamos primero que $p = q$, entonces se tiene:

$p_1 - p_0 = p_2 - p_1 = \cdots = p_{a+b} - p_{a+b-1} = c$

en donde c es una constante. Así que, para $z \in \{1, \ldots, a+b\}$:

$p_z - p_0 = \sum_{k=1}^{z} (p_k - p_{k-1}) = cz$

Por lo tanto, $p_z = p_0 + cz = 1 + cz$

Utilizando ahora la condición $p_{a+b} = 0$, se obtiene $c = -\frac{1}{a+b}$ y entonces:

$p_z = p_0 + cz = 1 - \frac{z}{a+b} = \frac{a+b-z}{a+b}$

En forma análoga, si, para $z \in \{0, \ldots, a+b\}$, denotamos por q_z a la probabilidad de que el jugador Q quede arruinado al comenzar P con z fichas, se obtiene $q_z = \frac{z}{a+b}$.

En particular, se tiene $p_z + q_z = 1$, es decir, con probabilidad 1 alguno de los dos jugadores se arruina.

Para el problema planteado inicialmente, en el cual el jugador P comienza con a fichas y el jugador Q con b fichas, la probabilidad de ruina del jugador P es entonces $\frac{b}{a+b}$.

Supongamos ahora $p \neq q$. En este caso tenemos, para $z \in \{1, \ldots, a+b-1\}$:

$q^z \prod_{k=1}^{z} (p_k - p_{k-1}) = p^z \prod_{k=1}^{z} (p_{k+1} - p_k)$

Así que:

$p_{z+1} - p_z = \left(\frac{q}{p}\right)^z (p_1 - p_0)$

y entonces, para $z \in \{1, \ldots, a+b\}$:

$p_z - p_0 = \sum_{k=0}^{z-1} (p_{k+1} - p_k) = (p_1 - p_0) \sum_{k=0}^{z-1} \left(\frac{q}{p}\right)^k = (p_1 - p_0)\, \frac{1-\left(\frac{q}{p}\right)^z}{1-\frac{q}{p}}$

Por lo tanto:

$p_z = 1 + (p_1 - 1)\, \frac{1-\left(\frac{q}{p}\right)^z}{1-\frac{q}{p}}$

Utilizando ahora la condición $p_{a+b} = 0$, se obtiene $p_1 - 1 = -\frac{1-\frac{q}{p}}{1-\left(\frac{q}{p}\right)^{a+b}}$ y entonces:

$p_z = 1 - \frac{1-\left(\frac{q}{p}\right)^z}{1-\left(\frac{q}{p}\right)^{a+b}} = \frac{\left(\frac{q}{p}\right)^z-\left(\frac{q}{p}\right)^{a+b}}{1-\left(\frac{q}{p}\right)^{a+b}}$

En forma análoga, si, para $z \in \{0, \ldots, a+b\}$, denotamos por q_z a la probabilidad de que el jugador Q quede arruinado al comenzar P con z fichas, se obtiene:

$q_z = \frac{\left(\frac{p}{q}\right)^{a+b-z}-\left(\frac{p}{q}\right)^{a+b}}{1-\left(\frac{p}{q}\right)^{a+b}} = \frac{1-\left(\frac{q}{p}\right)^z}{1-\left(\frac{q}{p}\right)^{a+b}}$

En particular, también en este caso se tiene $p_z + q_z = 1$. Es decir, con probabilidad 1, alguno de los dos jugadores se arruina.

Para el problema planteado inicialmente, en el cual el jugador P comienza con a fichas y el jugador Q con b fichas, la probabilidad de ruina del jugador P es entonces $\frac{\left(\frac{q}{p}\right)^a-\left(\frac{q}{p}\right)^{a+b}}{1-\left(\frac{q}{p}\right)^{a+b}}$.

Un caso interesante de este problema se obtiene al considerar un juego contra un jugador inmensamente rico. En otras palabras, supongamos que b es muy grande.

Si $p \leq q$, entonces, como es de esperarse, la probabilidad de ruina del jugador P tiende a 1 a medida que b crece. Pero si $p > q$ entonces, cuando b crece, la probabilidad de ruina del jugador P tiende a $\left(\frac{q}{p}\right)^a < 1$. En otras palabras, a pesar de tener una fortuna tan grande como se quiera, la probabilidad de ruina del jugador Q es positiva, incluso esta probabilidad podría ser cercana a 1. Por ejemplo si $\frac{q}{p} = \frac{1}{10}$, entonces bastaría con que P comenzara con una fortuna $a = 2$ para que ya la probabilidad de ruina de Q fuera muy cercana a $\frac{99}{100}$.

5.2. Probabilidades geométricas

Supongamos que tenemos una barra de extremos A y B y consideremos el siguiente experimento aleatorio: una bola W, la cual consideraremos sin dimensiones, es decir puntual, es lanzada al azar sobre la barra, de manera que la bola queda sobre ella. Si imaginamos la barra colocada sobre la recta real, de tal manera que los extremos de la barra correspondan a los puntos A y B sobre la recta, este experimento equivale a la elección al azar de un punto en el intervalo $[A, B]$. El espacio muestral Ω asociado a este experimento es entonces el mismo intervalo $[A, B]$. Ahora bien, si, como hemos considerado hasta este momento, cualquier subconjunto del espacio muestral representa un evento, entonces la construcción de un modelo matemático para este experimento consiste en la definición de una función de probabilidad P definida sobre la familia $\mathcal{A}$ formada por todos los subconjuntos de Ω. La definición de P deberá estar basada en el hecho de que **la elección del punto sobre el intervalo $[\mathbf{A}, \mathbf{B}]$**

se realiza al azar, lo cual podemos interpretar en el sentido de que dos subconjuntos de $[\mathbf{A}, \mathbf{B}]$ que se puedan sobreponer geométricamente uno sobre el otro deben de tener asignada la misma probabilidad.

Un problema del mismo tipo se puede plantear en el plano, en 3 dimensiones, o, de manera más general, en $\mathbb{R}^n$. Por ejemplo, en $\mathbb{R}^2$ consideremos el experimento aleatorio consistente en elegir al azar un punto al interior de un rectángulo, en $\mathbb{R}^3$ el de elegir al azar un punto al interior de un paralelepípedo, etc. Si interpretamos la elección al azar en el mismo sentido que antes, tendríamos que buscar entonces una función de probabilidad definida sobre todos los subconjuntos de la región en donde se elige el punto, de tal manera que dos subconjuntos que se puedan sobreponer geométricamente uno sobre el otro tengan asignada la misma probabilidad. Debido a la caracterización geométrica de lo que significa la elección al azar, este tipo probabilidades reciben el nombre de **probabilidades geométricas**.

El problema que se plantea en estos casos consiste en llevar a cabo el proceso de asignación de probabilidades a todos los subconjuntos de la región en donde se elige el punto al azar.

Consideremos, por ejemplo, el problema en una dimensión[5], es decir, supongamos que el experimento aleatorio consiste en la elección al azar de un punto dentro del intervalo $[A, B]$. Para cada punto $x \in [A, B]$, definamos el evento A_x: 'la bola cae en el punto x'.

Demostremos en primer lugar que la probabilidad del evento A_x es cero para cualquier x. En efecto, sea $\varepsilon > 0$ un número arbitrariamente pequeño y dividamos el segmento $\overline{AB}$ en N partes iguales de longitud menor que ε. Sean $A = x_0 < x_1 < \cdots < x_N = B$ los puntos de la subdivisión y, para $i \in \{0, \ldots, N-1\}$, definamos los siguientes eventos:

A_i: la bola cae en el intervalo $[x_i, x_{i+1})$.[6]

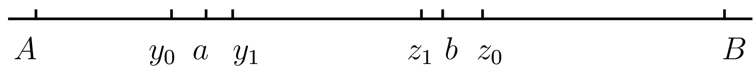
Los eventos A_i son mutuamente excluyentes y tienen una probabilidad común p pues se pueden sobreponer uno sobre el otro, de manera que tenemos $Np \leq 1$, es decir $p \leq \frac{1}{N} < \varepsilon$. Sea entonces A el evento A_i correspondiente al intervalo $[x_i, x_{i+1})$ que contiene a x. Evidentemente se tiene $P(A_x) \leq P(A) < \varepsilon$ y siendo ε arbitrariamente pequeña obtenemos $P(A_x) = 0$.

Sean ahora $a < b$ dos puntos cualesquiera en el intervalo $[A, B]$. Para cada $\varepsilon > 0$, arbitrariamente pequeña, consideremos una subdivisión del intervalo $[A, B]$ del tipo

[5] Los resultados para el caso de 2 o más dimensiones son similares.

[6] Si $x = B$, consideramos intervalos de la forma $(x_i, x_{i+1}]$.

descrito anteriormente, de tal manera que los puntos a y b queden en distintos subintervalos. Sean $[y_0, y_1)$ y $[z_1, z_0)$ los subintervalos en los cuales quedan comprendidos a y b respectivamente[7].



Consideremos entonces los siguientes eventos:

B: la bola cae en el intervalo (a, b).

C: la bola cae en el intervalo (y_1, z_1).

D: la bola cae en el intervalo (y_0, z_0).

B_i: la bola cae en el intervalo (x_i, x_{i+1}).

Evidentemente se tiene $P(C) \leq P(B) \leq P(D)$.

Ahora bien, los eventos B_i son mutuamente excluyentes y tienen una probabilidad común p. Además, la bola W cae en algún punto del intervalo $[A, B]$ con probabilidad 1 y se ha demostrado que $P(A_x) = 0$. Con base en esto se tiene:

$P(B_i) = \frac{1}{N} = \frac{x_{i+1}-x_i}{l}$

en donde l es la longitud del intervalo $[A, B]$.

Por otra parte, los intervalos (y_1, z_1) y (y_0, z_0) son unión de un número finito de puntos y de un número finito de intervalos (x_i, x_{i+1}), los cuales son ajenos; por lo tanto, tenemos también:

$P(C) = \frac{z_1-y_1}{l}$

$P(D) = \frac{z_0-y_0}{l}$

Así que:

$\frac{z_1-y_1}{l} \leq P(B) \leq \frac{z_0-y_0}{l}$

Pero como cada subintervalo tiene una longitud menor que ε, se tiene:

$z_1 - y_1 > (b - a) - 2\varepsilon$

$z_0 - y_0 < (b - a) + 2\varepsilon$

[7]Si $b = B$, consideramos intervalos de la forma $]x_i, x_{i+1}]$.

Entonces:

$\frac{b-a}{l} - \frac{2\varepsilon}{l} < P(B) < \frac{b-a}{l} + \frac{2\varepsilon}{l}$

Siendo ε arbitrariamente pequeña, se obtiene finalmente:

$P(B) = \frac{b-a}{l}$

Se concluye entonces que la condición de que la elección se realiza al azar determina de manera única la probabilidad de cada subintervalo del intervalo $[A, B]$, la cual es igual a la longitud de dicho subintervalo dividida por la longitud total del segmento en donde se elige el punto.

La pregunta que viene ahora es ¿hasta dónde se podrá continuar este proceso de asignación de probabilidades, de manera única, a los subconjuntos del intervalo $[A, B]$?

En el caso de elecciones al azar en el plano, se puede ver más claramente como llevar más lejos el proceso de asignación de probabilidades de manera única.

Sea R un rectángulo en el plano y consideremos el experimento aleatorio consistente en elegir al azar un punto al interior de R. En forma análoga al caso de una dimensión, se demuestra sin dificultad que la probabilidad de que el punto seleccionado sea un punto determinado o pertenezca a un segmento horizontal o vertical dentro de R, es igual a cero. También, la probabilidad de que pertenezca al interior de un rectángulo contenido en R es igual al área de dicho rectángulo dividida por el área de R.

Sea ahora S cualquier región contenida en R y sea A el evento 'el punto seleccionado pertenece a la región S'.

S

R

Partamos tanto la base como la altura de R en n partes iguales y sobre cada una de ellas tracemos líneas verticales y horizontales, respectivamente, que corten a R.

Esto determina una partición de R en rectángulos. Considerando que la función de probabilidad es finitamente aditiva, la probabilidad del evento A, en caso de que esté bien definida, queda acotada, por abajo, por la suma de las áreas de los rectángulos contenidos en la región S, dividida por el área de R, y por arriba, por la suma de las áreas de los rectángulos que contienen puntos de S, dividida por el área de R. Por lo tanto, cuando esas sumas tienden a un límite común cuando n tiende a ∞, ese límite común, dividido por el área de R, es la probabilidad del evento A. Se puede demostrar la coincidencia de los límites antes mencionados cuando S es cualquier región con un área bien definida con los métodos basados en la integral de Riemann que se estudian en los primeros cursos de Cálculo y, además, ese límite común es precisamente el área de S.

De esta manera, hemos llegado a la determinación de probabilidades geométricas en el plano para regiones S con un área bien definida: la probabilidad de que el punto seleccionado pertenezca a la región S es igual al cociente del área de S entre el área de R. Obsérvese que para tener este resultado es suficiente con que sea válido para el caso en que la región S es rectangular.

EJEMPLO 5.15 (**Problema del encuentro**). *Dos personas P y Q se citan en un determinado lugar entre las* 11 *y las* 12 *de la mañana, acordando que aquella que llegue primero esperará a la otra a lo más* 20 *minutos y en ningún caso después de la* 12*. Suponiendo que cada persona llega a la cita en un tiempo al azar entre las* 11 *y las* 12 *de manera que los tiempos de llegada son independientes uno del otro, calcule la probabilidad de que las dos personas se encuentren en el lugar de la cita.*

Solución

Los posibles tiempos de llegada a la cita los podemos representar por parejas de puntos (x, y), en donde x, y son los tiempos en que llegan P y Q, respectivamente. Si estos tiempos los contamos en minutos a partir de las 12*, los posibles tiempos de llegada quedarán representados por todos los puntos de un cuadrado de lado* 60 *sobre el plano.*

Sean $0 < a < b < 60$ *y* A *el evento 'P llega entre los tiempos* a *y* b*'. Por las hipótesis, se tiene:* $P(A) = \frac{b-a}{60}$.

Sea también $0 < c < d < 60$ *y* B *el evento 'Q llega entre los tiempos* c *y* d*'. Por las hipótesis, tenemos:* $P(B) = \frac{d-c}{60}$.

Ahora bien, siendo los tiempos de llegada independientes, tenemos: $P(A \cap B) = \frac{(b-a)(d-c)}{(60)^2}$.

Es decir, la probabilidad de que las llegadas de P *y* Q *estén comprendidas en el rectángulo* $(a, b) \times (c, d)$ *es igual al área del rectángulo dividida por el área total del cuadrado. Entonces, como dijimos antes, la probabilidad de un evento determinado por una región* S *dentro del cuadrado es igual al área de la región* S *dividida entre el área del cuadrado.*

Para que las dos personas se encuentren en el lugar de la cita es necesario y suficiente que $|x - y| \leq 20$*; por lo tanto el evento del cual deseamos calcular la probabilidad está representado por los puntos con esa propiedad, los cuales conforman la región* S *sombreada de la siguiente figura:*

El área de esta región es $(60)^2 - (40)^2$*; por lo tanto llamando* A *al evento 'P y Q se encuentran en el lugar de la cita', tenemos:*

$P(A) = \frac{(60)^2-(40)^2}{(60)^2} = 1 - \frac{4}{9} = \frac{5}{9}$

Ejemplo 5.16 (**Paradoja de Bertrand**). *Se traza una cuerda al azar en un círculo; ¿cuál es la probabilidad de que la longitud de la cuerda sea mayor que la longitud común de cada lado de un triángulo equilátero inscrito al círculo?*

Solución

La aparente paradoja proviene del hecho de que, al interpretar de diferentes maneras lo que significa trazar al azar una cuerda del círculo, se obtienen distintos valores de la probabilidad buscada. En efecto, resolveremos a continuación el problema planteado para algunas de esas interpretaciones. Siendo irrelevante la magnitud del radio del círculo, supondremos que éste es igual a 1.

a) Trazar una cuerda al azar puede interpretarse como equivalente a elegir dos puntos sobre el círculo de manera que las elecciones sean independientes y que, al elegir cada punto, hay la misma probabilidad de que éste esté comprendido en cualquiera de dos arcos de la misma longitud sobre el círculo.

La longitud de la cuerda excederá la longitud del lado del triángulo inscrito si y sólo si la longitud del arco que une A *con* B *en cualquier dirección es mayor que la tercera parte del perímetro del círculo. Entonces el problema es equivalente a elegir al azar dos puntos* x, y *del intervalo* $[0, 2\pi]$ *y queremos calcular la probabilidad de que se tenga* $\frac{4\pi}{3} \geq |y - x| \geq \frac{2\pi}{3}$ *. Esta probabilidad la podemos encontrar calculando el área de la región sombreada de la siguiente figura:*

Esta área es igual a $\left(\frac{4\pi}{3}\right)^2 - \left(\frac{2\pi}{3}\right)^2 = \frac{4\pi^2}{3}$*; por lo tanto, llamando* p *a la probabilidad buscada, tenemos* $p = \frac{\frac{4\pi^2}{3}}{(2\pi)^2} = \frac{1}{3}$.

b) Rotando el círculo después de elegir la cuerda, siempre podemos hacer que la cuerda quede en una dirección determinada de antemano, digamos en la dirección del eje x.

A

B

C

D

La cuerda elegida tendrá una longitud mayor a la longitud del triángulo equilátero inscrito siempre que caiga dentro de la región sombreada de la figura.

Ahora bien, la cuerda puede quedar determinada de varias maneras:

1o. Eligiendo al azar un punto sobre el semicírculo $ABCD$ y trazando por él una cuerda horizontal. En ese caso la cuerda queda en la región sombreada si y sólo si el punto se elige sobre el arco BC cuya longitud es la tercera parte del semicírculo, es decir, en este caso se obtiene $p = \frac{1}{3}$.

2o. Eligiendo al azar un punto sobre el intervalo $[-1, 1]$ en el eje y y trazando por él una cuerda horizontal, en ese caso la cuerda queda en la región sombreada si y sólo si el punto se elige sobre el intervalo $[-\frac{1}{2}, \frac{1}{2}]$, cuya longitud es la mitad de la de todo el intervalo, es decir, en este caso se obtiene $p = \frac{1}{2}$.

3o. Eligiendo al azar un punto dentro del círculo y trazando por él una cuerda horizontal, en este caso la cuerda queda en la región sombreada si y sólo si el punto se elige dentro de ésta, así que, siendo el área de la región sombreada $\frac{\pi}{3} + \frac{\sqrt{3}}{2}$, dividiendo esta cantidad entre el área total del círculo, obtendremos la probabilidad buscada; es decir $p = \frac{1}{3} + \frac{\sqrt{3}}{2\pi} \approx 0.69$.

5.3. Sucesiones infinitas de ensayos de Bernoulli

Consideremos el experimento aleatorio consistente en realizar una sucesión infinita de ensayos de Bernoulli, independientes uno de otro, en cada uno de los cuales la probabilidad de éxito es igual a p, en donde $0 < p \leq 1$.

Cada posible resultado de este experimento puede representarse mediante una sucesión infinita formada por éxitos y fracasos del tipo $(F, F, S, F, S, S, S, F, \ldots)$. Denotemos por ω_∞ a la sucesión formada exclusivamente por fracasos y por Ω^0 al conjunto de todas las otras sucesiones que componen el espacio muestral.

La ocurrencia del evento $\{\omega_\infty\}$ implica que, dado cualquier $n \in \mathbb{N}$, en los primeros n lanzamientos se obtiene fracaso, de manera que

$P(\{\omega_\infty\}) \leq (1-p)^n$, para toda $n \in \mathbb{N}$

Así que, $P(\{\omega_\infty\}) = 0$. Por lo tanto, $P(\Omega^0) = 1$. Lo cual significa que la probabilidad de obtener por lo menos un éxito es igual a 1.

Obsérvese que el resultado dice también que la probabilidad de obtener el primer éxito en un número finito de ensayos es igual a 1.

Al aplicar esta propiedad, aparentemente tan simple, se obtienen resultados que pueden parecer sorprendentes. Tomemos, por ejemplo, la obra Hamlet de Shakespeare y determinemos el conjunto C de caracteres tipográficos que ahí se utilizan (incluyendo el espacio blanco), así como el número total T de caracteres físicamente distintos que se utilizan en la obra. Supongamos ahora que una persona escribe una secuencia de T caracteres, cada uno seleccionado al azar del conjunto C. La probabilidad de que esa secuencia de T caracteres coincida exactamente con la obra de Shakespeare resulta ser igual a $p = \left(\frac{1}{m}\right)^T$, en donde m es el número de elementos que hay en C. Esta probabilidad p, aunque es extremadamente pequeña, es positiva. Supongamos ahora que el experimento consistente en escribir una secuencia de T caracteres, cada uno seleccionado al azar del conjunto C, se repite indefinidamente, siendo cada repetición independiente de las demás. En cada repetición de este experimento, definamos éxito al hecho de obtener, en esa repetición, una secuencia de T caracteres que coincida exactamente con la obra de Shakespeare. Tendremos entonces una sucesión de ensayos de Bernoulli, independientes, en cada uno de los cuales la probabilidad de éxito es igual a p. La relación $P(\Omega^0) = 1$ nos dice entonces que la probabilidad de que en algún momento una de las secuencias de T caracteres coincida exactamente con la obra de Shakespeare es igual a 1.

De igual forma, consideremos el experimento aleatorio consistente en realizar n ensayos de Bernoulli, independientes uno de otro, en cada uno de los cuales la probabilidad de éxito es igual a p, en donde $0 < p \leq 1$. La probabilidad de obtener éxito en todos los ensayos es entonces igual a $p^n > 0$.

Supongamos ahora que este experimento se repite indefinidamente, siendo cada repetición independiente de las demás. En cada repetición de este experimento, definamos éxito al hecho de obtener, en esa repetición, n éxitos. Tendremos entonces una sucesión de ensayos de Bernoulli, independientes, en cada uno de los cuales la probabilidad de éxito es igual a p^n. La relación $P(\Omega^0) = 1$ nos dice entonces que la probabilidad de que en algún momento se obtenga una secuencia de n éxitos es igual a 1. De esta manera, por ejemplo, no debe ser sorprendente el que en alguna ocasión una persona lance una moneda balanceada 100 veces y obtenga únicamente caras.

En general, sorprende que un evento de probabilidad muy pequeña ocurra. Sin embargo, por lo mencionado antes, la probabilidad de que tal evento ocurra por lo menos una vez en una sucesión de repeticiones independientes del correspondiente experimento aleatorio, es igual a 1. En contraparte, el número de ensayos que deban realizarse, en promedio, hasta obtener la ocurrencia de dicho evento, puede ser muy grande. Este punto se analizará con más detalle en el capítulo 9.

En realidad, cuando un resultado parece sorprendente es porque se está comparando un grupo de resultados con otro grupo distinto. Por ejemplo, la obtención de una secuencia de 100 caras al lanzar al aire 100 veces una moneda no debería ser sorprendente pues ese es uno de los posibles resultados que pueden obtenerse y todos ellos tienen la misma probabilidad de ocurrencia. Sin embargo dicho resultado puede parecer sorprendente al hacer diferentes tipos de comparaciones. Desde el punto de vista de la frecuencia de caras que se obtienen, es sorprendente pues el conjunto de posibles resultados con una frecuencia cercana a $\frac{1}{2}$ tiene una probabilidad bastante más alta que la correspondiente al conjunto de resultados cuya frecuencia difiere sustancialmente de $\frac{1}{2}$. Desde el punto de vista de la regularidad de la secuencia, es sorprendente pues el conjunto de posibles resultados con una regularidad visible es bastante más pequeño que los que no presentan una regularidad a simple vista, de manera que el primer grupo tiene una probabilidad más pequeña que el segundo.

De hecho, muchos métodos estadísticos están basados en las ideas expresadas arriba. Por ejemplo, para realizar una prueba de hipótesis se selecciona un conjunto de posibles resultados de tal manera que, bajo el supuesto de que la hipótesis es verdadera, su probabilidad es pequeña, y se adopta el criterio de rechazar la hipótesis si se obtiene un resultado de ese grupo. El criterio no garantiza que una hipótesis verdadera no va a ser rechazada pues, siendo verdadera es posible que ocurra uno de los posibles resultados del grupo que se está utilizando como criterio de rechazo, sin embargo, la pequeñez de la probabilidad de ese conjunto garantiza que la probabilidad de rechazar una hipótesis verdadera es pequeña. Hasta aquí cualquier conjunto de probabilidad pequeña es igualmente útil como criterio de rechazo, sin embargo se tiene que buscar también que la probabilidad de rechazar una hipótesis falsa sea alta. Los métodos para lograr esto pertenecen al campo de la Estadística Matemática.

Regresando al estudio de las sucesiones infinitas de ensayos de Bernoulli independientes, se puede decir todavía más. Para $m, k \in \mathbb{N}$, definamos los eventos:

A_m: Se obtienen menos de m éxitos.

B_k: En los primeros $k * m$ ensayos se obtienen menos de m éxitos.

Si se obtiene éxito en cada uno de los primeros m ensayos, entonces ocurre B_1^c, así que $p^m \leq P(B_1^c)$. Por lo tanto, $P(B_1) \leq 1 - p^m$. De la misma manera, para $k \geq 2$, $P(B_k \mid B_{k-1}) \leq 1 - p^m$, así que:

$$P(B_k) = P(B_k \mid B_{k-1})\, P(B_{k-1}) + P\left(B_k \mid B_{k-1}^c\right) P(B_{k-1}^c)$$

$$= P(B_k \mid B_{k-1})\, P(B_{k-1}) \leq (1 - p^m) P(B_{k-1})$$

Por lo tanto, $P(B_k) \leq (1 - p^m)^k$.

Ahora bien, si A_m ocurre, entonces ocurre B_k para cualquier k, así que $P(A_m) \leq (1 - p^m)^k$ para cualquier k. Por lo tanto, $P(A_m) = 0$. Lo cual significa que la probabilidad de obtener por lo menos m éxitos es igual a 1.

Si suponemos que la función de probabilidad es σ-aditiva, se puede enunciar un resultado todavía más fuerte. Definamos el evento:

A: se produce únicamente un número finito de éxitos.

Entonces, la sucesión de eventos $A_1, A_2, \ldots$ es monótona creciente y su unión es A. Se tiene entonces:

$$A = \bigcup_{m=1}^{\infty} A_m = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \cdots$$

Así que:

$$P(A) = P(A_1) + P(A_2 - A_1) + P(A_3 - A_2) + \cdots = 0$$

Por lo tanto, con probabilidad 1, se produce una infinidad de éxitos.

5.4. El problema de la medida

Consideremos el experimento aleatorio consistente en la elección al azar de un punto en el intervalo $[0, 1]$. Como dijimos antes, si cualquier subconjunto del espacio muestral representa un evento, la construcción de un modelo matemático para este experimento consiste en la definición de una función de probabilidad P definida sobre la familia $\mathcal{A}$ formada por todos los subconjuntos del intervalo $[0, 1]$. Ya vimos que la condición de que la elección se realiza al azar determina de manera única la probabilidad de cada subintervalo del intervalo $[0, 1]$, la cual es igual a la longitud de dicho subintervalo. Sin embargo, aún no tenemos el modelo completo pues no hemos asignado una probabilidad a cada subconjunto del intervalo $[0, 1]$. Por otra parte, la función de probabilidad que queremos definir debe de tener la propiedad de que dos subconjuntos del intervalo $[0, 1]$ que se puedan sobreponer geométricamente uno sobre el otro deben de tener asignada la misma probabilidad.

En el año 1902, Henri Léon Lebesgue ([**9**]) se planteó un problema similar al que queremos resolver. De manera específica, bautizó como **el problema de la medida** al consistente en encontrar una función m definida sobre todos los subconjuntos acotados de números reales y satisfaciendo las siguientes condiciones:

(i) m es no negativa.
(ii) m es σ-aditiva.
(iii) $m([0, 1]) = 1$
(iv) m es invariante bajo traslaciones.

Este problema fue atacado por diferentes personas, en particular por Giuseppe Vitali, quien demostró en 1905 que el problema de la medida de Lebesgue no tiene solución ([**13**]).

El resultado de Vitali provocó algo de insatisfacción pues de lo que se trata es de definir una función, llamada medida, que permita extender el concepto de longitud de un intervalo a todos los subconjuntos acotados de números reales, de tal manera que se satisfagan ciertas propiedades que razonablemente podrían esperarse para esta función. El resultado de Vitali planteó una disyuntiva, o bien se aceptan como razonables las condiciones i, ii, iii y iv y entonces se restringe la familia de conjuntos a los cuales se les puede asignar una medida, o bien se buscan condiciones menos restrictivas para la función m de tal manera que pueda definirse para cualquier subconjunto acotado de números reales.

En la búsqueda de alternativas para el problema de la medida, otros matemáticos lo estudiaron considerando diferentes modificaciones. Los resultados en este sentido son los siguientes:

Felix Hausdorf ([**5**]) propuso en 1914 el problema de la medida en sentido amplio, el cual consiste en encontrar una función m definida sobre todos los subconjuntos acotados de números reales y que satisfaga las siguientes condiciones:

(i) m es no negativa.
(ii) m es finitamente aditiva.
(iii) $m([0, 1]) = 1$
(iv) m es invariante bajo traslaciones.

Stefan Banach demostró en 1923 que este problema, así como el problema análogo en dos dimensiones, sí tiene solución ([**1**]). Sin embargo, el mismo Hausdorf había demostrado que el problema análogo en tres o más dimensiones no tiene solución.

Banach y Kazimierz Kuratowski se plantearon en 1929 el problema de encontrar una función m definida sobre todos los subconjuntos el intervalo $[0, 1]$ y satisfaciendo las siguientes condiciones:

(i) m es no negativa.
(ii) m es σ-aditiva.
(iii) Si I es un intervalo, entonces $m(I)$ es igual a la longitud de I.

El resultado de Banach y Kuratowski fue que tal problema no tiene solución ([**2**]). Esto sorprendió a muchos, por ejemplo a Paul Pierre Lévy quien pensaba que, al quitar a la medida la condición de ser invariante bajo traslaciones, es posible asignar una medida a todos los subconjuntos de números reales ([**10**]).

Alfred Tarski atacó en 1930 el problema de la medida en sentido amplio planteándose el problema de encontrar una función m definida sobre todos los subconjuntos del intervalo $[0, 1]$ y satisfaciendo las siguientes condiciones:

(i) m es no negativa.
(ii) m es finitamente aditiva.
(iii) Si I es un intervalo, entonces $m(I)$ es igual a la longitud de I.

Tarski mostró que tal problema, así como el análogo en dos o más dimensiones, sí tiene solución. Sin embargo, mostró también que la solución no es única ([**12**]).

Volviendo a nuestro problema de asignar una probabilidad a cada subconjunto del intervalo $[0, 1]$ de tal manera que dos subconjuntos del intervalo $[0, 1]$ que se puedan sobreponer geométricamente uno sobre el otro tengan asignada la misma probabilidad, vemos, a la luz de los resultados anteriores, que éste no tiene solución si se quiere una función de probabilidad que sea σ-aditiva, en cambio, si únicamente se quiere una función de probabilidad finitamente aditiva, el problema sí tiene solución, aunque, aún en ese caso, el problema análogo en tres dimensiones no lo tiene. De esta manera se puede concluir que la invariancia geométrica es muy restrictiva. Sustituyendo ésta por la condición de que un intervalo tenga asignada una probabilidad igual a su longitud, vemos que, si se quiere una función de probabilidad que sea σ-aditiva, el problema tampoco tiene solución. Por otra parte, si únicamente se quiere una función de probabilidad finitamente aditiva, el problema sí tiene solución, pero ésta no es única.

La conclusión que se deriva de estos resultados es que los experimentos descritos no pueden ser modelados dentro del marco que hemos considerado hasta este momento. Por lo tanto, se debe de tomar una de dos alternativas, o bien se renuncia a modelar este tipo de experimentos o bien se cambia alguna o algunas de las características del modelo.

Hemos visto que la modelación de experimentos consistentes en elecciones al azar, de los elementos de un conjunto finito, ha jugado un papel central para poder construir modelos probabilísticos en el caso general de experimentos aleatorios que admiten únicamente un conjunto finito de posibles resultados, e incluso para aquellos que admiten un conjunto infinito numerable. La misma situación se presenta en el caso de experimentos aleatorios que admiten una infinidad no numerable de posibles resultados, la modelación de experimentos consistentes en elecciones al azar juega también un papel central en la construcción de modelos probabilísticos en general. Además, se puede mostrar que el problema planteado para el caso de elecciones al azar se presenta en prácticamente todos los casos de experimentos aleatorios en donde el conjunto de posibles resultados es infinito no numerable. Por tal motivo, la renuncia a modelar tales tipos de experimentos empobrecería mucho la teoría desarrollada. De esta manera, la alternativa que queda consiste en relajar las características que se piden a la función de probabilidad.

Es posible tomar como modelo el que incluya una función de probabilidad que sea únicamente finitamente aditiva de tal manera que, para el experimento aleatorio planteado, se pueda asignar una probabilidad a cualquier subconjunto del intervalo $[0, 1]$. Con esta opción se estaría renunciando a la unicidad en la asignación de probabilidades. Esta opción es la que se toma, por ejemplo, en la Estadística Bayesiana. Sin embargo, en el enfoque clásico de la Teoría de la Probabilidad se busca que cada evento tenga asignada una probabilidad de manera única. Esto hace que la opción de tomar una función de probabilidad que sea únicamente finitamente aditiva tampoco resulte satisfactoria.

Abundando en el problema de la unicidad, se puede mostrar que si se quiere definir una función de probabilidad P σ-aditiva, esta función queda únicamente determinada sobre una familia $\mathcal{L}$ de subconjuntos del intervalo $[0, 1]$, la cual tiene las siguientes propiedades:

(i) Si I es cualquier subintervalo del intervalo $[0, 1]$, entonces $I \in \mathcal{L}$.
(ii) Si A es un subconjunto del intervalo $[0, 1]$ tal que $A \in \mathcal{L}$, entonces $[0, 1] - A \in \mathcal{L}$.
(iii) Si $A_1, A_2, \ldots$ es una colección finita, o infinita numerable, de elementos de $\mathcal{L}$, entonces $\bigcup_k A_k \in \mathcal{L}$.

Por otra parte, si se quiere definir una función de probabilidad P finitamente aditiva, se puede mostrar que esta función queda únicamente determinada sobre una familia $\mathcal{J}$ de subconjuntos del intervalo $[0, 1]$, la cual tiene las siguientes propiedades:

(i) Si I es cualquier subintervalo del intervalo $[0,1]$, entonces $I \in \mathcal{J}$.
(ii) Si A es un subconjunto del intervalo $[0,1]$ tal que $A \in \mathcal{J}$, entonces $[0,1]-A \in \mathcal{J}$.
(iii) Si $A_1, \ldots, A_n$ es una colección finita de elementos de $\mathcal{J}$, entonces $\bigcup_{k=1}^{n} A_k \in \mathcal{J}$.

Evidentemente la familia $\mathcal{L}$ contiene a la familia $\mathcal{J}$ pues una función de probabilidad σ-aditiva también es finitamente aditiva. Tal contención es propia. En efecto, se puede mostrar que hay subconjuntos del intervalo $[0,1]$ que pertenecen a $\mathcal{L}$ pero no a $\mathcal{J}$. Por ejemplo, la probabilidad de un subconjunto del intervalo $[0,1]$ que sea infinito numerable no queda determinada de manera única si únicamente se cuenta con la propiedad de la aditividad finita, en cambio, asumiendo que la función de probabilidad es σ-aditiva, la probabilidad de tal subconjunto queda únicamente determinada y es igual a cero pues los conjuntos formados por un punto tienen probabilidad igual a cero.

Vemos así que el proceso de asignación de probabilidades, de manera única, a los subconjuntos del espacio muestral se puede llevar más lejos con una función de probabilidad σ-aditiva que con una que sea simplemente finitamente aditiva. Esta es una de las razones básicas por las cuales se prefiere adoptar como modelo matemático para los experimentos aleatorios el que incluye la σ-aditividad de la función de probabilidad.

5.5. Espacios de Probabilidad

Lo expuesto en las secciones anteriores de este capítulo motiva las siguientes definiciones:

DEFINICIÓN 5.17 (**Álgebra de conjuntos**). *Sea Ω un conjunto. Se dice que una familia $\mathcal{A}$ de subconjuntos de Ω es un álgebra si se satisfacen las siguientes condiciones:*

(i) $\Omega \in \mathcal{A}$
(ii) *Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$.*
(iii) *Si $A_1, \ldots, A_n$ es cualquier familia finita de elementos de $\mathcal{A}$, entonces $\bigcup_{k=1}^{n} A_k \in \mathcal{A}$.*

DEFINICIÓN 5.18 (**$\boldsymbol{\sigma}$-álgebra**). *Sea Ω un conjunto. Se dice que una familia $\mathcal{A}$ de subconjuntos de Ω es una σ-álgebra si es un álgebra y dada cualquier colección numerable de elementos de $\mathcal{A}$, $A_1, A_2, \ldots$, entonces $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$.*

Definición 5.19 (**Función finitamente aditiva sobre un álgebra**). *Sea* Ω *un conjunto y* $\mathcal{A}$ *un álgebra de subconjuntos de* Ω*. Se dice que una función no negativa* $P : \mathcal{A} \mapsto \mathbb{R} \cup \{\infty\}$ *es finitamente aditiva si dada cualquier familia finita,* $A_1, \ldots, A_n$*, de elementos de* $\mathcal{A}$ *tal que* $A_i \cap A_j = \emptyset$ *para* $i \neq j$*, entonces* $P(\bigcup\limits_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$*.*

Definición 5.20 (**Función $\boldsymbol{\sigma}$-aditiva sobre un álgebra**). *Sea* Ω *un conjunto y* $\mathcal{A}$ *un álgebra de subconjuntos de* Ω*. Se dice que una función no negativa* $P : \mathcal{A} \mapsto \mathbb{R} \cup \{\infty\}$ *es* σ*-aditiva si es finitamente aditiva y dada cualquier familia numerable,* $A_1, A_2, \ldots$*, de elementos de* $\mathcal{A}$ *tal que* $A_i \cap A_j = \emptyset$ *para* $i \neq j$ *y* $\bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{A}$*, entonces* $P(\bigcup\limits_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$*.*

Definición 5.21 (**Función $\boldsymbol{\sigma}$-subaditiva**). *Sea* Ω *un conjunto y* $\mathcal{A}$ *un álgebra de subconjuntos de* Ω*. Se dice que una función no negativa* $P : \mathcal{A} \mapsto \mathbb{R} \cup \{\infty\}$ *es* σ*-subaditiva, o que satisface la propiedad de la subaditividad numerable, si dada cualquier familia finita o infinita numerable,* $A_1, A_2, \ldots$*, de elementos de* $\mathcal{A}$ *tal que* $\bigcup_k A_k \in \mathcal{A}$*,* $P(\bigcup_k A_k) \leq \sum_k P(A_k)$*.*

Debe de observarse que, en general, la σ-aditividad no es una consecuencia de la aditividad finita. Consideremos, por ejemplo, el álgebra $\mathcal{A}$ formada por los subconjuntos de los números naturales que son finitos o de complemento finito y definamos la función $P : \mathcal{A} \mapsto [0, 1]$ mediante las relaciones: $P(A) = 0$ si A es finito y $P(A) = 1$ si A^c es finito. Tal función es finitamente aditiva pero no σ-aditiva. Más aún, se puede mostrar ([**4**]) que P puede extenderse (no de manera única) a una función finitamente aditiva definida sobre la familia de todos los subconjuntos de los números naturales. Tal extensión, la cual está definida sobre una σ-álgebra, resulta entonces ser finitamente aditiva, pero no σ-aditiva.

Definición 5.22 (**Medida de probabilidad**). *Sea* Ω *un conjunto y* $\mathcal{A}$ *una* σ*-álgebra de subconjuntos de* Ω*. Se dice que una función* $P : \mathcal{A} \mapsto \mathbb{R}$ *es una medida de probabilidad si se satisfacen las siguientes propiedades:*

(i) $P(A) \geq 0$ *para todo evento* A*.*
(ii) $P(\Omega) = 1$
(iii) P *es* σ*-aditiva.*

De acuerdo con lo dicho en la sección anterior, un experimento aleatorio será modelado por una terna $(\Omega, \mathcal{A}, P)$, en donde Ω es un conjunto, $\mathcal{A}$ una σ-álgebra de subconjuntos de Ω y P una medida de probabilidad definida sobre $\mathcal{A}$. Una terna de este tipo será llamada un **espacio de probabilidad**. Este modelo está basado en la formulación axiomática de la Teoría de la Probabilidad dada por Andrey Nikolaevich Kolmogorov en el año 1933 ([**8**]). En su artículo, Kolmogorov mostró que la σ-aditividad admite un tratamiento matemático sólido, convenciendo así al medio matemático de aceptarla como una propiedad razonable del modelo probabilístico de cualquier experimento aleatorio.

El siguiente resultado muestra que la σ-aditividad se puede formular de diferentes maneras, todas ellas equivalentes.

PROPOSICIÓN 5.23. *Sea Ω un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de Ω y $P : \mathcal{A} \mapsto \mathbb{R}$ una función no negativa y finitamente aditiva, entonces las siguientes propiedades son equivalentes:*

(i) *P es σ-subaditiva.*
(ii) *P es σ-aditiva.*
(iii) *Para cualquier colección infinita $A_1, A_2, \ldots$ de elementos de $\mathcal{A}$ tales que $A_1 \subset A_2 \subset \ldots$ y $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, se tiene $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightsquigarrow \infty} P(A_n)$.*
(iv) *Para cualquier colección infinita $A_1, A_2, \ldots$ de elementos de $\mathcal{A}$ tales que $A_1 \supset A_2 \supset \ldots$ y $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$, se tiene $P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightsquigarrow \infty} P(A_n)$.*
(v) *Para cualquier colección infinita $A_1, A_2, \ldots$ de elementos de $\mathcal{A}$ tales que $A_1 \supset A_2 \supset \ldots$ y $\bigcap_{n=1}^{\infty} A_n = \emptyset$, se tiene $\lim_{n \rightsquigarrow \infty} P(A_n) = 0$.*

Demostración

$i \Longleftrightarrow ii$

Supongamos que P es σ-subaditiva y sea $A_1, A_2, \ldots$ una colección numerable de elementos de $\mathcal{A}$, ajenos por parejas y tales que $\bigcup_j A_j \in \mathcal{A}$. Se tiene:

$P(\bigcup_j A_j) \geq \sum_{k=1}^{n} P(A_j)$ para cualquier $n \in \mathbb{N}$, así que $P(\bigcup_j A_j) \geq \sum_{k=1}^{\infty} P(A_j)$.

Por lo tanto, P es σ-aditiva.

Supongamos ahora que P es σ-aditiva y sea $A_1, A_2, \ldots$ una colección numerable de elementos de $\mathcal{A}$ tales que $\bigcup_j A_j \in \mathcal{A}$. Sea $B_n = A_n - \bigcup_{j=1}^{n-1} A_j$, entonces $\bigcup_j A_j = \bigcup_j B_j$, así que:

$P(\bigcup_j A_j) = P(\bigcup_j B_j) = \sum_{k=1}^{\infty} P(B_j) \leq \sum_{k=1}^{\infty} P(A_j)$

$ii \Longrightarrow iii$

Supongamos que P es σ-aditiva y sea $A_1, A_2, \ldots$ una colección de elementos de $\mathcal{A}$ tales que $A_1 \subset A_2 \subset \ldots$ y $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, entonces:

$\bigcup_{n=1}^{\infty} A_n = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \cdots$

Así que:

$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{k=1}^{\infty} P(A_k - A_{k-1}) = \text{lím}_{n\rightsquigarrow\infty} \sum_{k=1}^{n} P(A_k - A_{k-1}) = \text{lím}_{n\rightsquigarrow\infty} P(A_n)$

en donde $A_0 = \emptyset$.

$iii \Longrightarrow iv$ es inmediato tomando complementos.

$iv \Longrightarrow v$ es inmediato pues v es un caso particular de iv.

$v \Longrightarrow ii$

Sea $A_1, A_2, \ldots$ una colección numerable de elementos de $\mathcal{A}$, ajenos por parejas y tales que $\bigcup_j A_j \in \mathcal{A}$. Sea $B_n = \bigcup_{j=1}^{\infty} A_j - \bigcup_{j=1}^{n} A_j$, entonces $B_1 \supset B_2 \supset \ldots$ y $\bigcap_{n=1}^{\infty} B_n = \emptyset$. Por lo tanto, $\text{lím}_{n\rightsquigarrow\infty} P(B_n) = 0$, así que:

$P(\bigcup_{j=1}^{\infty} A_j) = \text{lím}_{n\rightsquigarrow\infty} P(\bigcup_{j=1}^{n} A_j) = \text{lím}_{n\rightsquigarrow\infty} \sum_{k=1}^{n} P(A_j) = \sum_{k=1}^{\infty} P(A_j)$

∎

5.6. Teorema de clases monótonas

Si Ω es un conjunto arbitrario, se pueden definir distintas σ-álgebras de subconjuntos de Ω. Por ejemplo, si A es cualquier subconjunto de Ω, la familia $\mathcal{A} = \{\emptyset, \Omega, A, A^c\}$ constituye una σ-álgebra de subconjuntos de Ω. También la familia formada por todos los subconjuntos de Ω constituye una σ-álgebra.

Definición 5.24 (**Intersección de $\boldsymbol{\sigma}$-álgebras**). *Dado un conjunto Ω y una familia arbitraria de σ-álgebras de subconjuntos de Ω, se define la intersección de esas σ-álgebras como la familia de conjuntos que pertenecen a todas ellas.*

Se puede ver fácilmente que la intersección de σ-álgebras de subconjuntos de Ω es también una σ-álgebra y, dada una colección arbitraria $\mathcal{B}$ de subconjuntos de Ω, siempre existe por lo menos una σ-álgebra que contiene a todos los elementos de $\mathcal{B}$, a saber, la formada por todos los subconjuntos de Ω. Se puede definir entonces una σ-álgebra como la intersección de todas las σ-álgebras de subconjuntos de Ω que contienen a todos los elementos de $\mathcal{B}$. La σ-álgebra así definida es llamada la **σ-álgebra generada por $\mathcal{B}$** y evidentemente es la más pequeña σ-álgebra de subconjuntos de Ω que contiene a todos los elementos de $\mathcal{B}$.

Definición 5.25 (**$\boldsymbol{\sigma}$ álgebra generada por una familia de conjuntos**). *Dada una colección $\mathcal{A}$ de subconjuntos de un conjunto Ω, se define la σ-álgebra generada por $\mathcal{A}$ como la intersección de todas las σ-álgebras que contienen a todos los conjuntos de $\mathcal{A}$.*

Es frecuente encontrar un problema del siguiente tipo: Se tiene un conjunto Ω, un álgebra $\mathcal{A}$ de subconjuntos de Ω y la σ-álgebra $\sigma(\mathcal{A})$ generada por $\mathcal{A}$, y se quiere demostrar que una cierta propiedad es válida para todo elemento de $\sigma(\mathcal{A})$. Un método para resolver este problema consiste en demostrar que la familia $\mathcal{H}$ de subconjuntos de Ω que tienen la propiedad deseada es una σ-álgebra que contiene a todos los elementos de $\mathcal{A}$, de manera que entonces contiene a la σ-álgebra generada por $\mathcal{A}$. Sin embargo, en ocasiones puede resultar sumamente complicado demostrar que efectivamente $\mathcal{H}$ forma una σ-álgebra, sobre todo la demostración de que $\mathcal{H}$ es cerrada bajo uniones o intersecciones finitas. Por ejemplo para probar, siguiendo este método, que si dos medidas de probabilidad P_1 y P_2, definidas sobre $\sigma(\mathcal{A})$, coinciden sobre $\mathcal{A}$ entonces coinciden sobre $\sigma(\mathcal{A})$, se requeriría demostrar, en particular, que si A_1 y A_2 son dos elementos de $\sigma(\mathcal{A})$ tales que $P_1(A_1) = P_2(A_1)$ y $P_1(A_2) = P_2(A_2)$ entonces $P_1(A_1 \cup A_2) = P_2(A_1 \cup A_2)$ (o, equivalentemente, $P_1(A_1 \cap A_2) = P_2(A_1 \cap A_2)$), lo cual no hay manera de hacerlo, de manera directa, para dos medidas de probabilidad P_1 y P_2. Para salvar esta dificultad, se tienen, afortunadamente, los resultados que se exponen en esta sección, los cuales son conocidos como teoremas de clases monótonas.

Definición 5.26 (**Familia de conjuntos cerrada bajo diversas operaciones**). *Sea Ω un conjunto y $\mathcal{G}$ una familia de subconjuntos de Ω. Diremos que*

(i) *$\mathcal{G}$ es cerrada bajo complementos si $A^c \in \mathcal{G}$ para cualquier $A \in \mathcal{G}$.*
(ii) *$\mathcal{G}$ es cerrada bajo diferencias propias si $B - A \in \mathcal{G}$ para cualquier pareja $A, B \in \mathcal{G}$ tal que $A \subset B$.*
(iii) *$\mathcal{G}$ es cerrada bajo uniones (resp. intersecciones) finitas si $\bigcup_{j=1}^{n} A_j \in \mathcal{G}$ (resp. $\bigcap_{j=1}^{n} A_j \in \mathcal{G}$) para cualquier colección finita $A_1, \ldots, A_n$ de elementos de $\mathcal{G}$.*
(iv) *$\mathcal{G}$ es cerrada bajo uniones (resp. intersecciones) monótonas si $\bigcup_{j=1}^{\infty} A_j \in \mathcal{G}$ (resp. $\bigcap_{j=1}^{\infty} A_j \in \mathcal{G}$) para cualquier sucesión $A_1, A_2, \ldots$ de elementos de $\mathcal{G}$ tal que $A_1 \subset A_2 \subset \cdots$ (resp. $A_1 \supset A_2 \supset \cdots$).*

Definición 5.27 (**Clase monótona**). *Sea Ω un conjunto y $\mathcal{M}$ una familia de subconjuntos de Ω. Se dice que $\mathcal{M}$ es una clase monótona si es cerrada bajo uniones e intersecciones monótonas.*

Definición 5.28 (**Intersección de clases monótonas**). *Dado un conjunto Ω y una familia arbitraria de clases monótonas de subconjuntos de Ω, se define la intersección de esas clases monótonas como la familia de conjuntos que pertenecen a todas ellas.*

Se puede ver fácilmente que la intersección de clases monótonas, de subconjuntos de un conjunto Ω, forma una clase monótona.

Definición 5.29 (**clase monótona generada por una familia de conjuntos**). *Dada una colección $\mathcal{A}$ de subconjuntos de un conjunto Ω, se define la clase monótona generada por $\mathcal{A}$ como la intersección de todas las clases monótonas que contienen a todos los elementos de $\mathcal{A}$ y se le denota por $\mathcal{M}(\mathcal{A})$.*

Obsérvese que la definición anterior es consistente pues, dada cualquier colección $\mathcal{A}$ de subconjuntos de un conjunto Ω, existe por lo menos una clase monótona que contiene a todos los conjuntos de $\mathcal{A}$, a saber, la clase monótona formada por todos los subconjuntos de Ω.

Obsérvese también que la clase monótona generada por una familia $\mathcal{A}$ de subconjuntos de un conjunto Ω es la más pequeña clase monótona de subconjuntos de Ω que contiene a todos los elementos de $\mathcal{A}$.

TEOREMA 5.30. *Sea Ω un conjunto y $\mathcal{A}$ un álgebra de subconjuntos de Ω, entonces la clase monótona generada por $\mathcal{A}$ sigue siendo un álgebra.*

Demostración

Para demostrar que $\mathcal{M}(\mathcal{A})$ es cerrada bajo complementos, sea:

$\mathcal{C} = \{A \in \mathcal{M}(\mathcal{A}) : A^c \in \mathcal{M}(\mathcal{A})\}$

$\mathcal{C}$ es entonces una clase monótona. En efecto, si $A_1 \subset A_2 \cdots$ es una sucesión de elementos de $\mathcal{C}$ entonces, para cualquier $n \in \mathbb{N}$, se tiene $A_n \in \mathcal{M}(\mathcal{A})$ y $A_n^c \in \mathcal{M}(\mathcal{A})$, así que $\bigcup_n A_n \in \mathcal{M}(\mathcal{A})$ y $(\bigcup_n A_n)^c = \bigcap_n A_n^c \in \mathcal{M}(\mathcal{A})$, por lo tanto, $\bigcup_n A_n \in \mathcal{C}$. De la misma manera se demuestra que $\mathcal{C}$ es cerrada bajo intersecciones monótonas.

Obviamente $\mathcal{C}$ contiene a $\mathcal{A}$, de manera que entonces se puede concluir que $\mathcal{C}$ contiene a $\mathcal{M}(\mathcal{A})$, lo cual prueba que $\mathcal{M}(\mathcal{A})$ es cerrada bajo complementos.

Para demostrar que $\mathcal{M}(\mathcal{A})$ es cerrada bajo intersecciones finitas, sea:

$\mathcal{C}_1 = \{A \in \mathcal{M}(\mathcal{A}) : A \cap B \in \mathcal{M}(\mathcal{A})$ para cualquier $B \in \mathcal{A}\}$

$\mathcal{C}_1$ es entonces una clase monótona. En efecto, si $A_1 \subset A_2 \cdots$ es una sucesión de elementos de $\mathcal{C}_1$ entonces, para cualquier $n \in \mathbb{N}$ y $B \in \mathcal{A}$, se tiene $A_n \in \mathcal{M}(\mathcal{A})$ y $A_n \cap B \in \mathcal{M}(\mathcal{A})$, así que $\bigcup_n A_n \in \mathcal{M}(\mathcal{A})$ y $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B) \in \mathcal{M}(\mathcal{A})$, por lo tanto, $\bigcup_n A_n \in \mathcal{C}_1$. De la misma manera se demuestra que $\mathcal{C}_1$ es cerrada bajo intersecciones monótonas.

Obviamente $\mathcal{C}_1$ contiene a $\mathcal{A}$, de manera que entonces se puede concluir que $\mathcal{C}_1$ contiene a $\mathcal{M}(\mathcal{A})$, es decir, para cualquier $A \in \mathcal{M}(\mathcal{A})$ y $B \in \mathcal{A}$, se tiene $A \cap B \in \mathcal{M}(\mathcal{A})$.

Sea ahora:

$\mathcal{C}_2 = \{A \in \mathcal{M}(\mathcal{A}) : A \cap B \in \mathcal{M}(\mathcal{A})$ para cualquier $B \in \mathcal{M}(\mathcal{A})\}$

$\mathcal{C}_2$ es entonces una clase monótona. En efecto, si $A_1 \subset A_2 \cdots$ es una sucesión de elementos de $\mathcal{C}_2$ entonces, para cualquier $n \in \mathbb{N}$ y $B \in \mathcal{M}(\mathcal{A})$, se tiene $A_n \in \mathcal{M}(\mathcal{A})$ y $A_n \cap B \in \mathcal{M}(\mathcal{A})$, así que $\bigcup_n A_n \in \mathcal{M}(\mathcal{A})$ y $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B) \in \mathcal{M}(\mathcal{A})$, por lo tanto, $\bigcup_n A_n \in \mathcal{C}_2$. De la misma manera se demuestra que $\mathcal{C}_2$ es cerrada bajo intersecciones monótonas.

Como $\mathcal{C}_1$ contiene a $\mathcal{M}(\mathcal{A})$, se tiene que $\mathcal{C}_2$ contiene a $\mathcal{A}$, de manera que entonces se puede concluir que $\mathcal{C}_2$ contiene a $\mathcal{M}(\mathcal{A})$, es decir, para cualesquiera $A, B \in \mathcal{M}(\mathcal{A})$, se tiene $A \cap B \in \mathcal{M}(\mathcal{A})$.

$\blacksquare$

COROLARIO 5.31 (**Teorema de clases monótonas para álgebras**). *Sea Ω un conjunto y $\mathcal{A}$ un álgebra de subconjuntos de Ω, entonces $\mathcal{M}(\mathcal{A}) = \sigma(A)$.*

PROPOSICIÓN 5.32. *Sea Ω un conjunto, $\mathcal{A}$ un álgebra de subconjuntos de Ω y P_1 y P_2 dos medidas de probabilidad, definidas sobre $\sigma(\mathcal{A})$, tales que $P_1(A) = P_2(A)$ para cualquier $A \in \mathcal{A}$, entonces $P_1(A) = P_2(A)$ para cualquier $A \in \sigma(A)$.*

Demostración

Sea $\mathcal{H} = \{A \in \sigma(\mathcal{A}) : P_1(A) = P_2(A)\}$

$\mathcal{H}$ es entonces una clase monótona que contiene a los elementos de $\mathcal{A}$, así que $\mathcal{H}$ contiene a $\sigma(\mathcal{A}) = \mathcal{M}(\mathcal{A})$.

$\blacksquare$

El teorema 5.30 puede ser insuficiente pues en ocasiones únicamente se puede demostrar inmediatamente que la propiedad que se quiere probar como válida para cualquier elemento de una σ-álgebra se cumple para una familia de conjuntos que la generan y que es cerrada bajo intersecciones finitas (por ejemplo, la familia formada por todos los intervalos de números reales).

DEFINICIÓN 5.33 (**$\boldsymbol{\pi}$ sistema**). *Sea Ω un conjunto y $\mathcal{P}$ una familia de subconjuntos de Ω. Se dice que $\mathcal{P}$ es un π-sistema si es cerrada bajo intersecciones finitas.*

DEFINICIÓN 5.34 (**d-sistema**). *Sea Ω un conjunto y $\mathcal{D}$ una familia de subconjuntos de Ω. Se dice que $\mathcal{D}$ es un d-sistema si $\Omega \in \mathcal{D}$ y es cerrada bajo diferencias propias y uniones monótonas.*

DEFINICIÓN 5.35 (**Intersección de d-sistemas**). *Dado un conjunto Ω y una familia arbitraria de d-sistemas de subconjuntos de Ω, se define la intersección de esos d-sistemas como la familia de conjuntos que pertenecen a todos ellos.*

Se puede ver fácilmente que la intersección de d-sistemas, de subconjuntos de un conjunto Ω, forma un d-sistema.

DEFINICIÓN 5.36 (**d-sistema generado por una familia de conjuntos**). *Dada una colección $\mathcal{G}$ de subconjuntos de un conjunto Ω, se define el d-sistema generado por*

$\mathcal{G}$ como la intersección de todos los d-sistemas que contienen a todos los elementos de $\mathcal{G}$ y se le denota por $d(\mathcal{G})$.

Obsérvese que la definición anterior es consistente pues, dada cualquier colección $\mathcal{P}$ de subconjuntos de un conjunto Ω, existe por lo menos un d-sistema que contiene a todos los conjuntos de $\mathcal{P}$, a saber, el d-sistema formada por todos los subconjuntos de Ω.

Obsérvese también que el d-sistema generado por una familia $\mathcal{P}$ de subconjuntos de un conjunto Ω es el más pequeño d-sistema de subconjuntos de Ω que contiene a todos los elementos de $\mathcal{P}$.

Teorema 5.37. *Sea Ω un conjunto y $\mathcal{P}$ un π-sistema de subconjuntos de Ω, entonces el d-sistema generado por $\mathcal{P}$ es un π-sistema.*

Demostración

Sea:

$\mathcal{C}_1 = \{A \in d(\mathcal{P}) : A \cap B \in d(\mathcal{P})$ para cualquier $B \in \mathcal{P}\}$

$\mathcal{C}_1$ es entonces un d-sistema. En efecto, si $A_1 \subset A_2 \cdots$ es una sucesión de elementos de $\mathcal{C}_1$, entonces, para cualquier $n \in \mathbb{N}$ y $B \in \mathcal{P}$, se tiene $A_n \in d(\mathcal{P})$ y $A_n \cap B \in d(\mathcal{P})$, así que $\bigcup_n A_n \in d(\mathcal{P})$ y $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B) \in d(\mathcal{P})$; por lo tanto, $\bigcup_n A_n \in \mathcal{C}_1$. Sean ahora $A, C \in \mathcal{C}_1$ tales que $A \subset C$, entonces $A, C, A \cap B, C \cap B \in d(\mathcal{P})$ y $A \cap B \subset C \cap B$ para cualquier $B \in \mathcal{P}$, así que $C - A \in d(\mathcal{P})$ y $(C - A) \cap B = C \cap B - A \cap B \in d(\mathcal{P})$, por lo tanto, $C - A \in \mathcal{C}_1$. Finalmente, es obvio que $\Omega \in \mathcal{C}_1$.

Obviamente $\mathcal{C}_1$ contiene a $\mathcal{P}$, de manera que entonces se puede concluir que $\mathcal{C}_1$ contiene a $d(\mathcal{P})$, es decir, para cualquier $A \in d(\mathcal{P})$ y $B \in \mathcal{P}$, se tiene $A \cap B \in d(\mathcal{P})$.

Sea ahora:

$\mathcal{C}_2 = \{A \in d(\mathcal{P}) : A \cap B \in d(\mathcal{P})$ para cualquier $B \in d(\mathcal{P})\}$

$\mathcal{C}_2$ es entonces un d-sistema. En efecto, si $A_1 \subset A_2 \cdots$ es una sucesión de elementos de $\mathcal{C}_2$ entonces, para cualquier $n \in \mathbb{N}$ y $B \in d(\mathcal{P})$, se tiene $A_n \in d(\mathcal{P})$ y $A_n \cap B \in d(\mathcal{P})$, así que $\bigcup_n A_n \in d(\mathcal{P})$ y $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B) \in d(\mathcal{P})$; por lo tanto, $\bigcup_n A_n \in \mathcal{C}_2$. Sean ahora $A, C \in \mathcal{C}_2$ tales que $A \subset C$, entonces $A, C, A \cap B, C \cap B \in d(\mathcal{P})$ y $A \cap B \subset C \cap B$ para cualquier $B \in d(\mathcal{P})$, así que $C - A \in d(\mathcal{P})$ y $(C - A) \cap B = C \cap B - A \cap B \in d(\mathcal{P})$, por lo tanto, $C - A \in \mathcal{C}_2$. Finalmente, es obvio que $\Omega \in \mathcal{C}_2$.

Como $\mathcal{C}_1$ contiene a $d(\mathcal{P})$, se tiene que $\mathcal{C}_2$ contiene a $\mathcal{P}$, de manera que entonces se puede concluir que $\mathcal{C}_2$ contiene a $d(\mathcal{P})$, es decir, para cualesquiera $A, B \in d(\mathcal{P})$, se tiene $A \cap B \in d(\mathcal{P})$.

■

Corolario 5.38 (**Teorema de clases monótonas para π-sistemas**). *Sea Ω un conjunto y $\mathcal{P}$ un π-sistema de subconjuntos de Ω, entonces $d(\mathcal{P}) = \sigma(\mathcal{P})$.*

Proposición 5.39. *Sea Ω un conjunto, $\mathcal{P}$ un π-sistema de subconjuntos de Ω tal que $\Omega \in \mathcal{P}$ y P_1 y P_2 dos medidas de probabilidad, definidas sobre $\sigma(\mathcal{P})$, tales que $P_1(A) = P_2(A)$ para cualquier $A \in \mathcal{P}$, entonces $P_1(A) = P_2(A)$ para cualquier $A \in \sigma(\mathcal{P})$.*

Demostración

Sea $\mathcal{H} = \{A \in \sigma(\mathcal{P}) : P_1(A) = P_2(A)\}$.

$\mathcal{H}$ es entonces un d-sistema que contiene a los elementos de $\mathcal{P}$, así que $\mathcal{H}$ contiene a $\sigma(\mathcal{P}) = d(\mathcal{P})$.

■

Las proposiciones 5.32 y 5.39 son de gran importancia para la Teoría de la Probabilidad pues muestran que **el proceso de extensión de una función de probabilidad, definida sobre un álgebra o un π-sistema, a una medida de probabilidad sobre la σ-álgebra que generan, en caso de poder realizarse, se puede lograr de una única manera**. Es decir, la extensión es única.

5.7. Los borelianos y la medida de Lebesgue

El problema de la medida planteado por H. Lebesgue consiste en extender el concepto de longitud a una familia tan grande como sea posible de subconjuntos de números reales. De manera más específica, se trata de comenzar asignando a cada intervalo su longitud y después de extender esa función tan lejos como sea posible. Como ya lo mencionamos, tal extensión se puede realizar hasta abarcar una $\boldsymbol{\sigma}$-álgebra de subconjuntos de $\mathbb{R}$, la generada por los intervalos.

Definición 5.40 (**$\boldsymbol{\sigma}$-álgebra de Borel en $\mathbb{R}$**). *La σ-álgebra de Borel en $\mathbb{R}$ es la σ-álgebra de subconjuntos de $\mathbb{R}$ generada por la familia de intervalos de la forma $(-\infty, x]$, en donde $x \in \mathbb{R}$. A los elementos de esa σ-álgebra se les llama borelianos de $\mathbb{R}$.*

Las siguientes proposiciones pueden ser probadas fácilmente.

Proposición 5.41. *Todo intervalo de $\mathbb{R}$ es boreliano.*

Proposición 5.42. *Los subconjuntos abiertos y los subconjuntos cerrados de $\mathbb{R}$ son borelianos.*

Proposición 5.43. *La σ-álgebra de los conjuntos borelianos de $\mathbb{R}$ está generada por cualquiera de las siguientes familias de subconjuntos.*

a) Los intervalos de la forma (x, ∞).

b) Los intervalos de la forma $(-\infty, x)$.

c) Los intervalos de la forma $[x, \infty)$.

d) Los intervalos de la forma $(a, b]$.

e) Los intervalos de la forma $[a, b)$.

f) Los intervalos de la forma (a, b).

g) Los intervalos de la forma $[a, b]$.

h) Los conjuntos abiertos.

i) Los conjuntos cerrados.

Definición 5.44 (**$\boldsymbol{\sigma}$-álgebra de Borel en $(\mathbf{a}, \mathbf{b})$**). *La σ-álgebra de Borel en el intervalo (a, b) es la σ-álgebra de subconjuntos de (a, b) formada por las intersecciones de los borelianos en $\mathbb{R}$ con el intervalo (a, b). A los elementos de esa σ-álgebra se les llama borelianos de (a, b).*

Definición 5.45 (**Medida cero**). *Se dice que un conjunto $A \subset \mathbb{R}$ tiene medida cero si, para cualquier $\varepsilon > 0$, existe una colección finita o infinita numerable de intervalos abiertos $\{I_n\}$ tales que $A \subset \bigcup_n I_n$ y $\sum_n l(I_n) < \varepsilon$, en donde $l(I_n)$ denota la longitud del intervalo I_n.*

El siguiente resultado se demuestra fácilmente:

Proposición 5.46. *Si $\{A_n\}$ es una familia finita o infinita numerable de conjuntos de medida cero, entonces $A = \cup_n A_n$ tiene medida cero.*

Ejemplo 5.47. *Cualquier conjunto numerable $A \subset \mathbb{R}$ tiene medida cero. En efecto, si $A = \{x_1, x_2, \ldots\}$, dada $\varepsilon > 0$, consideremos, para cada $n \in \mathbb{N}$, un intervalo abierto I_n, de centro x_n y longitud $\frac{\varepsilon}{2^{n+1}}$, entonces $A \subset \bigcup_n I_n$ y $\sum_n l(I_n) = \sum_{n=1}^{\infty} \frac{\varepsilon}{2^{n+1}} = \varepsilon \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} = \frac{\varepsilon}{2} < \varepsilon$.*

▲

La familia de conjuntos borelianos, si bien es bastante grande, presenta una limitación consistente en que un conjunto de medida cero puede no ser boreliano. En un problema de probabilidad, esto a su vez se traduce en que puede haber subconjuntos de conjuntos de probabilidad cero para los cuales, por no ser borelianos, no está definida la función de probabilidad. Por esa razón, es conveniente considerar una σ-álgebra más grande que los borelianos, que incluya a todos los conjuntos de medida cero. A continuación se muestra cómo puede hacerse tal extensión.

LEMA 5.48. *Todo conjunto de medida cero está contenido en un conjunto boreliano de medida cero.*

Demostración

Sea $A \subset (0,1)$ un conjunto de medida cero, entonces, para cualquier $\varepsilon > 0$, existe una colección finita o infinita numerable de intervalos abiertos $\{I_k\}$ tales que $A \subset \bigcup_k I_k$ y $\sum_k l(I_k) < \varepsilon$, en donde $l(I_k)$ denota la longitud del intervalo I_k.

Para cada $n \in \mathbb{N}$, sea $\{I_k^n\}$ una colección finita o infinita numerable de intervalos abiertos $\{I_k^n\}$ tales que $A \subset \bigcup_k I_k^n$ y $\sum_k l(I_k^n) < \frac{1}{n}$. Definamos entonces $B_n = \bigcup_k I_k^n$ y $B = \bigcap_n B_n$. Entonces, B es boreliano, tiene medida cero y $A \subset B$.

■

DEFINICIÓN 5.49 (**Conjunto Lebesgue medible**). *Diremos que un conjunto $A \subset \mathbb{R}$ es Lebesgue medible si pertenece a la σ-álgebra de subconjuntos de $\mathbb{R}$ generada por los borelianos y los conjuntos de medida cero.*

PROPOSICIÓN 5.50. *Todo conjunto Lebesgue medible se puede expresar como la unión de un conjunto boreliano y un conjunto de medida cero.*

Demostración

Denotemos por $\mathcal{B}$ a los conjuntos borelianos y sea:

$\mathcal{H} = \{H \subset \mathbb{R} : H = B \cup C$, en donde $B \in \mathcal{B}$ y C tiene medida cero$\}$

Obviamente, cualquier elemento de $\mathcal{H}$ es Lebesgue medible y $\mathcal{H}$ contiene tanto a los borelianos como a los conjuntos de medida cero. Para probar el resultado, basta entonces con demostrar que $\mathcal{H}$ es una σ-álgebra.

$\mathcal{H}$ contiene a $\mathbb{R}$ y es cerrada bajo uniones numerables. La única dificultad consiste en demostrar que $\mathcal{H}$ es cerrada bajo complementos.

Sea $H \in \mathcal{H}$, entonces H puede expresarse en la forma $H = B \cup C$, en donde B es un conjunto boreliano y C es un conjunto de medida cero.

Por el lema 5.48, existe un conjunto boreliano D de medida cero tal que $C \subset D$. Sea $A = D \cap B^c$, entonces:

$H^c = (B \cup C)^c = (B \cup A)^c \bigcup (A - C)$

Así que $H^c \in \mathcal{H}$.

■

Corolario 5.51. *La σ-álgebra de los conjuntos Lebesgue medibles está formada por todos los conjuntos de la forma $B \cup C$, en donde B es un conjunto boreliano y C un conjunto de medida cero.*

Proposición 5.52. *Sea $\mathcal{L}$ la σ-álgebra de los conjuntos Lebesgue medibles de $\mathbb{R}$ y $m_1 : \mathcal{L} \mapsto \mathbb{R} \cup \{\infty\}$, $m_2 : \mathcal{L} \mapsto \mathbb{R} \cup \{\infty\}$ dos funciones no negativas y σ-aditivas tales que $m_1(I) = m_2(I) =$ longitud(I) para cualquier intervalo I, entonces $m_1(B) = m_2(B)$ para cualquier $B \in \mathcal{L}$.*

Demostración

Por la proposición 5.39, m_1 y m_2 coinciden sobre los conjuntos borelianos contenidos en cualquier intervalo finito I.

Sea B un conjunto boreliano de $\mathbb{R}$, entonces:

$B = \bigcup_{n \in \mathbb{Z}} B \cap [n, n+1)$

Así que:

$m_1(B) = \sum_{n \in \mathbb{Z}} m_1 \left(B \cap [n, n+1)\right)$

$= \sum_{n \in \mathbb{Z}} m_2 \left(B \cap [n, n+1)\right) = m_2(B)$

Sea ahora A un conjunto de medida cero, entonces, para cualquier $\varepsilon > 0$, existe una colección finita o infinita numerable de intervalos abiertos $\{I_n\}$ tales que $A \subset \bigcup_n I_n$ y $\sum_n l(I_n) < \varepsilon$, en donde $l(I_n)$ denota la longitud del intervalo I_n. Por lo tanto:

$m_1(A) \leq m_1 \left(\bigcup_n I_n\right) \leq \sum_n m_1 \left(I_n\right) = \sum_n l \left(I_n\right) < \varepsilon$

Así que, $m_1(A) = 0$.

De la misma manera, se tiene, $m_2(A) = 0$. Por lo tanto, m_1 y m_2 coinciden sobre los conjuntos de medida cero . Así que, nuevamente por la proposición 5.39, m_1 y m_2 coinciden sobre los conjuntos Lebesgue medibles contenidos en cualquier intervalo finito I.

Sea $B \in \mathcal{L}$, entonces:

$B = \bigcup_{n \in \mathbb{Z}} B \cap [n, n+1)$

Así que:

$m_1(B) = \sum_{n \in \mathbb{Z}} m_1 \left(B \cap [n, n+1)\right)$

$= \sum_{n\in\mathbb{Z}} m_2 \left(B \cap [n, n+1)\right) = m_2(B)$ ■

DEFINICIÓN 5.53 (**Medida de Lebesgue en** $\mathbb{R}$). *La medida de Lebesgue m en $\mathbb{R}$ es la única función no negativa y σ-aditiva definida sobre los conjuntos Lebesgue medibles tal que $m(I)$ es igual a la longitud de I para cualquier intervalo I*[8].

EJEMPLO 5.54. *Demuestre que el conjunto de Cantor $\mathcal{C}$ es Lebesgue medible y que $m(\mathcal{C}) = 0$.*

Solución

$[0,1] - \mathcal{C} = \left(\frac{1}{3}, \frac{2}{3}\right) \cup \left(\frac{1}{3^2}, \frac{2}{3^2}\right) \cup \left(\frac{7}{3^2}, \frac{8}{3^2}\right) \cup \left(\frac{1}{3^3}, \frac{2}{3^3}\right) \cup \left(\frac{7}{3^3}, \frac{8}{3^3}\right) \cup \left(\frac{19}{3^3}, \frac{20}{3^3}\right) \cup \left(\frac{25}{3^3}, \frac{26}{3^3}\right) \cup \cdots$

Así que $\mathcal{C}$ es Lebesgue medible y:

$m(\mathcal{C}) = 1 - m\left([0,1] - \mathcal{C}\right)$

$= 1 - m\left(\left(\frac{1}{3}, \frac{2}{3}\right)\right) - m\left(\left(\frac{1}{3^2}, \frac{2}{3^2}\right)\right) - m\left(\left(\frac{7}{3^2}, \frac{8}{3^2}\right)\right) - \cdots$

$= 1 - \left(\frac{1}{3} + \frac{2}{3^2} + \frac{2^2}{3^3} + \frac{2^3}{3^4} + \cdots\right) = 1 - \frac{\frac{1}{3}}{\frac{1}{3} - \frac{2}{3}} = 0$

▲

Las definiciones y resultados para el caso de $\mathbb{R}$ pueden extenderse a $\mathbb{R}^n$, como se enuncia a continuación.

Por un rectángulo en $\mathbb{R}^n$ se entenderá un conjunto de la forma $I_1 \times \cdots \times I_n$, en donde $I_1, \cdots, I_n$ son intervalos en $\mathbb{R}$. Además, diremos que el rectángulo $R = I_1 \times \cdots \times I_n$ es abierto (resp. cerrado) si los intervalos $I_1, \cdots, I_n$ son abiertos (resp. cerrados).

DEFINICIÓN 5.55 (**$\boldsymbol{\sigma}$-álgebra de Borel en** $\mathbb{R}^n$). *La σ-álgebra de Borel en $\mathbb{R}^n$ es la σ-álgebra de subconjuntos de $\mathbb{R}^n$ generada por la familia de los rectángulos en $\mathbb{R}^n$ de la forma $(-\infty, x_1] \times \cdots \times (-\infty, x_n]$. A los elementos de esa σ-álgebra se les llama borelianos de $\mathbb{R}^n$.*

PROPOSICIÓN 5.56. *Todo rectángulo de $\mathbb{R}^n$ es boreliano.*

PROPOSICIÓN 5.57. *Los subconjuntos abiertos y los subconjuntos cerrados de $\mathbb{R}^n$ son borelianos.*

PROPOSICIÓN 5.58. *La σ-álgebra de los conjuntos borelianos de $\mathbb{R}^n$ está generada por cualquiera de las siguientes familias de subconjuntos:*

(i) *Los rectángulos abiertos.*
(ii) *Los rectángulos cerrados.*

[8] La existencia de tal medida fue demostrada por Henri Léon Lebesgue en el año 1902 ([**9**]).

(iii) *Los conjuntos abiertos.*
(iv) *Los conjuntos cerrados.*

DEFINICIÓN 5.59 (**σ-álgebra de Borel en un boreliano**). *La σ-álgebra de Borel en el boreliano $A \subset \mathbb{R}^n$ es la σ-álgebra de subconjuntos de A formada por las intersecciones de los borelianos en $\mathbb{R}^n$ con A. A los elementos de esa σ-álgebra se les llama borelianos de A.*

DEFINICIÓN 5.60 (**Medida cero**). *Se dice que un conjunto $B \subset \mathbb{R}^n$ tiene medida cero si, para cualquier $\varepsilon > 0$, existe una colección finita o infinita numerable de rectángulos $R_k = I_1^k \times \cdots \times I_n^k$ tales que $B \subset \bigcup_k R_k$ y $\sum_k l(I_1^k) \cdots l(I_n^k) < \varepsilon$, en donde $l(I)$ denota la longitud del intervalo I.*

PROPOSICIÓN 5.61. *Si $\{B_k\}$ es una familia finita o infinita numerable de conjuntos de medida cero, entonces $A = \bigcup_k B_k$ tiene medida cero.*

DEFINICIÓN 5.62 (**Conjunto Lebesgue medible en $\mathbb{R}^n$**). *Diremos que un conjunto $B \subset \mathbb{R}^n$ es Lebesgue medible si pertenece a la σ-álgebra de subconjuntos de $\mathbb{R}^n$ generada por los borelianos y los conjuntos de medida cero.*

PROPOSICIÓN 5.63. *Todo conjunto Lebesgue medible se puede expresar como la unión de un conjunto boreliano y un conjunto de medida cero.*

COROLARIO 5.64. *La σ-álgebra de los conjuntos Lebesgue medibles está formada por todos los conjuntos de la forma $B \cup C$, en donde B es un conjunto boreliano y C un conjunto de medida cero.*

PROPOSICIÓN 5.65. *Sea $\mathcal{L}^n$ la σ-álgebra de los conjuntos Lebesgue medibles de $\mathbb{R}^n$ y $m_1 : \mathcal{L}^n \mapsto \mathbb{R} \cup \{\infty\}$, $m_2 : \mathcal{L}^n \mapsto \mathbb{R} \cup \{\infty\}$ dos funciones no negativas y σ-aditivas tales que $m_1(R) = m_2(R) = l(I_1) \cdots l(I_n)$ para cualquier rectángulo $R = I_1 \times \cdots \times I_n$, en donde $l(I)$ denota la longitud del intervalo I, entonces $m_1(B) = m_2(B)$ para cualquier $B \in \mathcal{L}^n$.*

DEFINICIÓN 5.66 (**Medida de Lebesgue en $\mathbb{R}^n$**). *La medida de Lebesgue m en $\mathbb{R}^n$ es la única función no negativa y σ-aditiva definida sobre los conjuntos Lebesgue medibles de $\mathbb{R}^n$ tal que $m(R) = l(I_1) \cdots l(I_n)$ para cualquier rectángulo $R = I_1 \times \cdots \times I_n$, en donde $l(I)$ denota la longitud del intervalo I.*

PROPOSICIÓN 5.67. *Si $B_1, \ldots, B_n$ son borelianos de $\mathbb{R}$, entonces $B_1 \times B_2 \times \ldots \times B_n$ es un boreliano de $\mathbb{R}^n$.*

Demostración

Sean $x_1, \ldots, x_{n-1}$ números reales cualesquiera, entonces la familia de conjuntos $B \subset \mathbb{R}$ tales que $(-\infty, x_1] \times \cdots \times (-\infty, x_{n-1}] \times B$ es un boreliano de $\mathbb{R}^n$, forma una σ-álgebra que contiene a los intervalos de la forma $(-\infty, x]$; por lo tanto, contiene a todos los borelianos de $\mathbb{R}$.

De la misma manera, si $B_n \subset \mathbb{R}$ es un boreliano cualquiera, entonces la familia de conjuntos $B \subset \mathbb{R}$ tales que $(-\infty, x_1] \times \cdots \times (-\infty, x_{n-2}] \times B \times B_n$ es un boreliano de $\mathbb{R}^n$, forma una σ-álgebra que contiene a los intervalos de la forma $(-\infty, x]$; por lo tanto, contiene a todos los borelianos de $\mathbb{R}$.

Continuando con este procedimiento, se obtiene que los conjuntos de la forma $(-\infty, x_1] \times B_2 \times \ldots \times B_n$, en donde x_1 es un número real cualquiera y $B_2, \ldots, B_n$ son borelianos cualesquiera de $\mathbb{R}$, son borelianos de $\mathbb{R}^n$.

Finalmente, si $B_2, \ldots, B_n$ son borelianos cualesquiera de $\mathbb{R}$, entonces la familia de conjuntos $B \subset \mathbb{R}$ tales que $B \times B_2 \times \ldots \times B_n$ es un boreliano de $\mathbb{R}^n$, forma una σ-álgebra que contiene a los intervalos de la forma $(-\infty, x]$; por lo tanto, contiene a todos los borelianos de $\mathbb{R}$.

■

5.8. Funciones borelianas

DEFINICIÓN 5.68 (**Función boreliana**). *Una función $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ es llamada boreliana si $f^{-1}(B)$ es un conjunto boreliano de $\mathbb{R}^n$ para todo boreliano B de $\mathbb{R}^m$.*

Recuérdese que si f es cualquier función de $\mathbb{R}^n$ en $\mathbb{R}^m$ y $B, B_1, B_2, \ldots$ son subconjuntos de $\mathbb{R}^m$, entonces $f^{-1}(B^c) = [f^{-1}(B)]^c$, $f^{-1}(\bigcup_n B_n) = \bigcup_n f^{-1}(B_n)$ y $f^{-1}(\bigcap_n B_n) = \bigcap_n f^{-1}(B_n)$. Esto implica que la familia de subconjuntos B de $\mathbb{R}^m$, tales que $f^{-1}(B)$ es un conjunto boreliano de $\mathbb{R}^n$, forman una σ-álgebra de subconjuntos de $\mathbb{R}^m$, de manera que, para demostrar que una cierta función f es boreliana, basta con probar que $f^{-1}(B)$ es un conjunto boreliano de $\mathbb{R}^n$ para cualquier elemento B de una familia de generadores de los borelianos en $\mathbb{R}^m$. Con base en esta idea, se pueden demostrar las siguientes proposiciones, algunas de cuyas demostraciones se dejan como ejercicio:

PROPOSICIÓN 5.69. *Si $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ y $g : \mathbb{R}^m \mapsto \mathbb{R}^p$ son borelianas, entonces $g \circ f$ es boreliana.*

PROPOSICIÓN 5.70. *Toda función continua $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ es boreliana.*

PROPOSICIÓN 5.71. *Sean f y g funciones de $\mathbb{R}$ en $\mathbb{R}$ y $H : \mathbb{R} \mapsto \mathbb{R}^2$ definida por $H(x) = (f(x), g(x))$, entonces H es boreliana si y sólo si f y g son borelianas.*

Demostración

Supongamos que f y g son borelianas, entonces, $f^{-1}((-\infty, x])$ y $g^{-1}((-\infty, y])$ son conjuntos borelianos de $\mathbb{R}$ para cualesquiera $x, y \in \mathbb{R}$. Por lo tanto:

$$H^{-1}((-\infty, x] \times (-\infty, y]) = f^{-1}((-\infty, x]) \bigcap g^{-1}((-\infty, y])$$

es un boreliano de $\mathbb{R}$.

Por otra parte, la familia de subconjuntos $B \subset \mathbb{R}^2$ tales que $H^{-1}(B)$ es un boreliano de $\mathbb{R}$ forma una σ-álgebra de subconjuntos de $\mathbb{R}^2$ la cual, como se demostró antes, contiene a los rectángulos de la forma $(-\infty, x] \times (-\infty, y]$; por lo tanto, contiene a todos los borelianos de $\mathbb{R}^2$.

Inversamente, supongamos que la función $H : \mathbb{R} \mapsto \mathbb{R}^2$ definida por $H(x) = (f(x), g(x))$ es boreliana, entonces:

$f^{-1}\left((-\infty, x]\right) = H^{-1}\left((-\infty, x] \times \mathbb{R}\right)$

y:

$g^{-1}\left((-\infty, y]\right) = H^{-1}\left(\mathbb{R} \times (-\infty, y]\right)$

son borelianos de $\mathbb{R}$ para cualesquiera $x, y \in \mathbb{R}$.

Finalmente, la familia de subconjuntos $B \subset \mathbb{R}$ tales que $f^{-1}(B)$ (resp. $g^{-1}(B)$) es un boreliano de $\mathbb{R}$ forma una σ-álgebra de subconjuntos de $\mathbb{R}$ la cual, como se demostró antes, contiene a los rectángulos de la forma $(-\infty, x]$ (resp. $(-\infty, y]$); por lo tanto, contiene a todos los borelianos de $\mathbb{R}$. Es decir, f (resp. g) es boreliana.

■

Corolario 5.72. *Si $f : \mathbb{R} \mapsto \mathbb{R}$ y $g : \mathbb{R} \mapsto \mathbb{R}$ son borelianas y $a, b, \alpha \in \mathbb{R}$ con $\alpha \geq 0$, entonces $af + b$, $|f|^\alpha$, $f + g$, fg, $\text{máx}(f, g)$ y $\text{mín}(f, g)$ son borelianas.*

Proposición 5.73. *Sea $g : \mathbb{R}^n \mapsto \mathbb{R}$ una función boreliana no negativa. Entonces existe una sucesión monótona no decreciente de funciones borelianas no negativas φ_m tales que:*

(i) *Cada φ_m es de la forma:*
$\varphi_m = \sum_{j=0}^{k_m} b_{m,j} I_{B_{m,j}}$
en donde $b_{m,0}, \ldots, b_{m,k_m}$ son números reales positivos.
y $B_{m,0}, \ldots, B_{m,k_m}$ son conjuntos borelianos de $\mathbb{R}^n$.

(ii) $\text{lím}_{m \rightsquigarrow \infty} \varphi_m(x) = g(x)$ *para cualquier $x \in \mathbb{R}^n$.*

Demostración

Para cada $m \in \mathbb{N}$, sea:

$$\varphi_m(x) = \begin{cases} \sum_{k=1}^{m2^m} \frac{k-1}{2^m} I_{\left[\frac{k-1}{2^m} \leq g < \frac{k}{2^m}\right]} & \text{si } g(x) < m \\ 0 & \text{si } g(x) \geq m \end{cases}$$

Dados $m \in \mathbb{N}$ y $x \in \mathbb{R}^n$, sea k el único número natural tal que $\frac{k-1}{2^m} \leq g(x) < \frac{k}{2^m}$. Entonces, como $\varphi_m(x) = \frac{k-1}{2^m}$, se tiene $g(x) - \frac{1}{2^m} < \varphi_m(x) \leq g(x)$, así que $\lim_{m \rightsquigarrow \infty} \varphi_m(x) = g(x)$.

Ahora bien, como $\frac{2(k-1)}{2^{m+1}} \leq g(x) < \frac{2k}{2^{m+1}}$, se tiene que $\frac{2k-2}{2^{m+1}} \leq g(x) < \frac{2k-1}{2^{m+1}}$ o bien $\frac{2k-1}{2^{m+1}} \leq g(x) < \frac{2k}{2^{m+1}}$. En el primer caso, se tiene $\varphi_{m+1}(x) = \frac{2k-2}{2^{m+1}} = \frac{k-1}{2^m} = \varphi_m(x)$ mientras que en el segundo, se tiene $\varphi_{m+1}(x) = \frac{2k-1}{2^{m+1}} > \frac{2k-2}{2^{m+1}} = \varphi_m(x)$. Así que, en cualquier caso, $\varphi_m(\omega) \leq \varphi_{m+1}(\omega)$.

Así que, φ_m es una sucesión monótona no decreciente de funciones borelianas no negativas tales que $\lim_{m \rightsquigarrow \infty} \varphi_m(x) = g(x)$ para cualquier $x \in \mathbb{R}^n$.

■

El siguiente resultado será utilizado en el capítulo 9. Se trata de una propiedad que permite realizar el vínculo entre la Teoría de la Medida y la Teoría de Integración con un enfoque geométrico. Cuando uno quiere calcular el área de la región comprendida bajo la gráfica de una función no negativa, se integra tal función. En otras palabras, la integral de Riemann de una función no negativa corresponde al área bajo la gráfica de esa función. Lo mismo ocurre con la integral definida por H. Lebesgue. Para que este resultado adquiera sentido se requiere que, dada una función boreliana no negativa, la región comprendida bajo la gráfica de la función sea un conjunto al cual se le puede asignar área, es decir, que sea Lebesgue medible. Como se muestra a continuación, tal región es un conjunto boreliano de $\mathbb{R}^2$.

Proposición 5.74. *Sea $g : \mathbb{R}^n \mapsto \mathbb{R}$ cualquier función boreliana no negativa. Entonces la función $h : \mathbb{R}^{n+1} \mapsto \mathbb{R}$ definida por $h(x, y) = I_{[0,g(x))}(y)$, en donde $x \in \mathbb{R}^n$ y $y \in \mathbb{R}$, es boreliana.*

Demostración

Para cada $m \in \mathbb{N}$, sea:

$$\varphi_m(x) = \begin{cases} \sum_{k=0}^{m2^m} \frac{k-1}{2^m} I_{\left[\frac{k-1}{2^m} \leq g < \frac{k}{2^m}\right]} & \text{si } g(x) < m \\ 0 & \text{si } g(x) \geq m \end{cases}$$

De acuerdo con la demostración de la proposición anterior, φ_m es una sucesión monótona no decreciente de funciones no negativas tales que $\lim_{m \rightsquigarrow \infty} \varphi_m(x) = g(x)$ para cualquier $x \in \mathbb{R}^n$. Por lo tanto:

$$\{(x, y) \in \mathbb{R}^{n+1} : 0 \leq y < g(x)\} = \cup_{m=1}^{\infty} \{(x, y) \in \mathbb{R}^{n+1} : 0 \leq y < \varphi_m(x)\}$$

$$= \cup_{m=1}^{\infty} \left(\cup_{k=0}^{m2^m} \left[\left[\frac{k-1}{2^m} \leq g < \frac{k}{2^m}\right] \times \left[0, \frac{k-1}{2^m}\right)\right]\right)$$

Así que el conjunto $\{(x, y) \in \mathbb{R}^{n+1} : h(x, y) = 1\}$ es boreliano.

■

COROLARIO 5.75. *Sea $g : \mathbb{R}^n \mapsto \mathbb{R}$ una función boreliana no negativa. Entonces los siguientes conjuntos son borelianos de $\mathbb{R}^{n+1}$:*

$\{(x_1, \dots, x_n, y) \in \mathbb{R}^{n+1} : 0 < y < g(x_1, \dots, x_n)\}$

$\{(x_1, \dots, x_n, y) \in \mathbb{R}^{n+1} : 0 \leq y < g(x_1, \dots, x_n)\}$

$\{(x_1, \dots, x_n, y) \in \mathbb{R}^{n+1} : 0 < y \leq g(x_1, \dots, x_n)\}$

$\{(x_1, \dots, x_n, y) \in \mathbb{R}^{n+1} : 0 \leq y \leq g(x_1, \dots, x_n)\}$

$\{(x_1, \dots, x_n, y) \in \mathbb{R}^{n+1} : g(x) > 0 \text{ y } 0 \leq y \leq g(x_1, \dots, x_n)\}$

EJERCICIOS

EJERCICIO 5.1. *Resuelva los ejemplos 5.12 y 5.13 siguiendo el método de Huygens, es decir, utilizando probabilidades condicionales.*

EJERCICIO 5.2. *Muestre que los siguientes problemas caen dentro de la categoría de problemas descritos en la proposición 5.9 y encuentre las probabilidades que se piden.*

a) Dos jugadores, A y B, juegan a lanzar sucesivamente un par de dados, con la condición de que A ganará el juego en el momento en que obtenga 6 puntos, mientras que B lo ganará en el momento en que obtenga 7 puntos. A tiene el primer turno y después, comenzando con B, cada uno de ellos tendrá dos oportunidades consecutivas de ganar. ¿Qué probabilidad tiene cada jugador de ganar el juego.[9]

b) Tres jugadores, A, B y C, juegan un juego consistente en seleccionar, por turnos, una ficha, al azar y con reemplazo, de una caja que contiene 4 fichas blancas y 8 negras. Si el ganador es el primero que obtenga una ficha blanca, ¿qué probabilidad tiene cada jugador de ganar el juego?[10]

EJERCICIO 5.3. *Demuestre que en el problema de la ruina del jugador, si $a + b > 3$, entonces Ω^I no es numerable.*

EJERCICIO 5.4. *Demuestre que en el problema de la ruina del jugador, $\sum_{\{\omega \in \Omega^F\}} p(\omega) = 1$.*

[9]Este problema fue propuesto a C. Huygens por P. Fermat en junio de 1656.

[10]Este problema fue propuesto por C. Huygens en su libro ([**7**]).

EJERCICIO 5.5 (**Problema de los N jugadores**). *N jugadores, $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_N$, compiten por parejas de tal manera que en cada competencia cada uno de ellos tiene una probabilidad de ganar igual a $\frac{1}{2}$. Primero A_1 compite contra A_2 y el ganador compite contra A_3; el nuevo ganador compite contra A_4, etc. El ganador del juego es el primer jugador que logre vencer a los otros $N-1$ jugadores de manera consecutiva.*

El experimento aleatorio $\mathcal{E}$ asociado con este problema puede verse como la realización de una sucesión de ensayos de Bernoulli $B_1, B_2, \ldots$, en donde B_1 es éxito si el jugador P gana la primera partida y, para $i \geq 1$, B_{i+1} es éxito si el jugador que ganó la i-ésima partida gana también la que sigue. Con estas convenciones, el juego se termina en el momento en que se obtienen, por primera vez, $N-2$ éxitos después del primer ensayo.

El espacio muestral Ω de $\mathcal{E}$ puede descomponerse en dos subconjuntos Ω^F y Ω^I, en donde Ω^F es el conjunto de resultados de las sucesiones finitas de ensayos de Bernoulli que terminan con la primera ocurrencia de $N-2$ éxitos consecutivos después del primer ensayo y Ω^I es el conjunto de resultados de las sucesiones infinitas de ensayos de Bernoulli para las cuales nunca se obtienen $N-2$ éxitos consecutivos después del primer ensayo. Como antes, cada elemento de Ω^F puede representarse mediante una sucesión finita $(s_1, s_2, \ldots, s_n)$ de 0's y 1's y cada elemento de Ω^I puede representarse mediante una sucesión infinita $(s_1, s_2, \ldots)$ de 0's y 1's, en donde $s_i = 0$ y $s_i = 1$ representan, respectivamente, un fracaso y un éxito en el i-ésimo ensayo. Además, si $\omega = (s_1, \ldots, s_n)$ es una colección finita de 0's y 1's, definimos $p(\omega) = \frac{1}{2^n}$.

Demuestre que:

a) Si $N \geq 4$, entonces Ω^I no es numerable.

b) $\sum_{\omega \in \Omega^F} p(\omega) = 1$

EJERCICIO 5.6. *Dos personas, A y B, quedan de verse en un determinado lugar a las 12 hrs. Cada una de ellas llega al lugar de la cita en un tiempo al azar entre las 12 y las 12 : 30 hrs. Una vez que llega al lugar de la cita, la persona A está dispuesta a esperar a lo más 5 minutos a que llegue la persona B, mientras que la persona B está dispuesta a esperar a la persona A a lo más 10 minutos. ¿Cuál es la probabilidad de que las 2 personas se encuentren?*

EJERCICIO 5.7. *En una parada, el primer autobús del día pasa en un tiempo al azar entre las 7 : 00 y las 7 : 15 hrs. Si una persona llega a la parada en un tiempo al azar entre las 7 : 10 y las 7 : 15 hrs., ¿cuál es la probabilidad de que la persona llegue a la parada antes que el primer autobús del día?*

EJERCICIO 5.8. *Para ir de una parte de la ciudad a otra, una persona puede tomar cualquiera de dos transportes, los cuales siguen rutas distintas, identificadas con las* *letras A y B. Los transportes que siguen la ruta A salen cada 15 minutos comenzando a las 7 de la mañana, los que siguen la ruta B salen también cada 15 minutos pero comienzan a la $7:05$ de la mañana. La persona en consideración llega siempre al inicio de las rutas en un tiempo al azar entre las 7 y las 8 de la mañana y toma el primer transporte que salga después de su llegada. ¿Cuál es el porcentaje de veces en que la persona toma el transporte que sigue la ruta A?*

EJERCICIO 5.9. *Sea $A_1, A_2, \ldots$ una colección de eventos tales que $P(A_n) = 1$ para cualquier $n \in \mathbb{N}$. Demuestre que $P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1$.*

EJERCICIO 5.10. *Sean $A_1, A_2, \ldots$ y $B_1, B_2, \ldots$ dos colecciones de eventos tales que:*

$\text{lím}_{n \rightsquigarrow \infty} P(A_n) = 1$ *y* $\text{lím}_{n \rightsquigarrow \infty} P(B_n) = p$.

Demuestre que $\text{lím}_{n \rightsquigarrow \infty} P(A_n \cap B_n) = p$.

EJERCICIO 5.11. *Demuestre las proposiciones 5.41, 5.42 y 5.43.*

EJERCICIO 5.12. *Demuestre la proposición 5.46.*

EJERCICIO 5.13. *Demuestre las proposiciones 5.69 y 5.70.*

EJERCICIO 5.14. *Demuestre el corolario 5.72.*

Referencias

[1] Banach, S., Sur le problème de la mesure, Fundamenta Mathematicae, 4, p. 7-33, 1923.

[2] Banach, S. & Kuratowski, C., Sur une généralisation du problème de la mesure, Fundamenta Mathematicae, t. 14, p. 127-131, 1929.

[3] Bernoulli, J., L'Art de Conjecturer, L.G.F. Vastel, G. Le Roy, Caen, 1801. Traducción de Ars Conjectandi, Basileae, 1713.

[4] Bhaskara Rao, K.P.S. and Bhaskara Rao, M., Theory of Charges (A study of finitely additive measure), Academic Press, 1983.

[5] Hausdorff, F., Grundzüge der Mengenlehre, Chelsea Publishing Company, 1914.

[6] Horn, A., and Tarski, A., Measures in boolean algebras, Trans. Amer. Math. Soc., t. 64, p. 467-497, 1948.

[7] Huygens, C., Du calcul dans les jeux de hasard, Oeuvres Complètes de Christiaan Huygens, Vol. XIV, Martinus Nijhoff, 1920. Traducción de De Ratiociniis in Aleae Ludo, 1657.

[8] Kolmogorov, A. N., Foundations of the Theory of Probability, Chelsea, 1950. Traducción de Grundbegriffe der Wahrscheinlichkeitsrechnung, Erg Mat. 2, No. 3, 1933.

[9] Lebesgue, H. L., Leçons sur l'intégration et la recherche des fonctions primitives, Gauthier-Villars, 1904.

[10] Lévy, P. P., Les lois de probabilité dans les ensembles abstraits, Revue de Métaphysique et Morale, 1924. Reproducido en Calcul des Probabilités, Gauthier Villars, 1925.

[11] Los, J. & Marczewski, E., Extensions of measures, Fundamenta Mathematicae, t. 36, p. 267-276, 1949.

[12] Tarski, A., Une contribution a la Théorie de la Mesure, Fundamenta Mathematicae, t. 15, p. 42-50, 1930.

[13] Vitali, G., Sul problema della misura dei gruppi di punti di una retta, Tip. Bamberini et Parmeggiani, Bologna, 1905.

Parte 2

VARIABLES ALEATORIAS

CAPÍTULO 6

VARIABLES ALEATORIAS

La Matemática es el arte de dar el mismo nombre a diferentes cosas. La Poesía es el arte de dar diferentes nombres a la misma cosa.

Jules Henri Poincaré

En la ciencia uno trata de decir a la gente, de una manera que sea entendido por cualquiera, algo que nadie nunca conoció antes. Pero en la poesía, es exactamente lo opuesto.

Paul Dirac

Es cierto que un matemático que no tiene algo de poeta nunca será un perfecto matemático.

Karl Theodor Wilhelm Weierstrass

En todo este capítulo se asume que se tiene un espacio de probabilidad $(\Omega, \Im, P)$ correspondiente a un determinado experimento aleatorio.

Dados los conjuntos $\Omega, E, E_1, \ldots, E_n$, las funciones $X : \Omega \mapsto E, X_1 : \Omega \mapsto E_1, \ldots, X_n : \Omega \mapsto E_n$ y los conjuntos $B \subset E$, $B_1 \subset E_1$, ..., $B_n \subset E_n$, denotaremos por $[X \in B]$ al conjunto $\{\omega \in \Omega : X(\omega) \in B\}$ y por $[X_1 \in B_1, \ldots, X_n \in B_n]$ a la intersección de los conjuntos $\{\omega \in \Omega : X_k(\omega) \in B_k\}$, para $k \in \{1, \ldots, n\}$. También, si $A \subset E_1 \times \ldots \times E_n$, denotaremos por $[(X_1, \ldots, X_n) \in A]$ al conjunto $\{\omega \in \Omega : (X_1(\omega), \ldots, X_n(\omega)) \in A\}$.

En el caso en que B sea un intervalo de la forma $[a, b]$, también se utilizará la notación $[a \leq X \leq b]$ para el conjunto $[X \in B]$, con la desigualdad estricta si el intervalo es abierto en el extremo correspondiente. Para los intervalos infinitos se utilizará una notación similar, por ejemplo, si $B = (-\infty, b]$, denotaremos por $[X \leq b]$ al conjunto $[X \in B]$.

En lo que sigue se hará referencia a los conjuntos y funciones borelianas, conceptos que fueron estudiados en la sección 5.7.

6.1. Variables aleatorias reales

En ocasiones, dado un experimento aleatorio, estamos interesados en alguna característica numérica del resultado que se obtiene, por ejemplo, al considerar el lanzamiento de un par de dados, pudiéramos estar interesados no en el resultado específico que se obtiene con cada dado sino en la suma de ellos. El valor numérico que se obtiene depende del resultado específico del experimento, de manera que la característica numérica de interés puede representarse mediante una función $X : \Omega \mapsto \mathbb{R}$. Ahora bien, como el valor de X depende del resultado del experimento, antes de realizar éste, no se conoce su valor, pero, como se tiene definida una función de probabilidad, dado un conjunto B de números reales es posible calcular la probabilidad del conjunto $[X \in B]$ siempre que éste represente un evento. Es por esto que no toda función $X : \Omega \mapsto \mathbb{R}$ se considera una variable aleatoria, para que lo sea se requiere que los conjuntos $[X \in B]$ representen eventos para una familia suficientemente grande de subconjuntos B de números reales.

Definición 6.1 (**Variable aleatoria real**). *Se dice que una función $X : \Omega \mapsto \mathbb{R}$ es una variable aleatoria real si $[X \leq x] \in \Im$ para cualquier $x \in \mathbb{R}$.*

Ejemplo 6.2. *En los siguientes casos, X es una variable aleatoria:*

(i) *Un experimento aleatorio consiste en seleccionar al azar una bola de una urna, la cual contiene N bolas numeradas del 1 al N, y X es el número de la bola seleccionada.*
(ii) *Un experimento aleatorio consiste en tomar una muestra aleatoria con reemplazo de tamaño n de una urna, la cual contiene r bolas rojas y b bolas blancas, y X es el número de bolas blancas en la muestra.*
(iii) *Un experimento aleatorio consiste en lanzar 3 dados y X es la diferencia entre el mayor y el menor de los números que se obtienen.*
(iv) *Un experimento aleatorio consiste en lanzar una moneda indefinidamente y X es el número de caras que se obtienen antes de la primera cruz.*
(v) *Un experimento aleatorio consiste en elegir al azar un número del intervalo $(0, 1)$ y X es el número que se obtiene.*

(vi) *Un experimento aleatorio consiste en elegir al azar un número del intervalo* $(0, N)$ *y* X *es la parte entera del número que se obtiene.*

▲

Obsérvese que si A es un evento, entonces I_A es una variable aleatoria y, en el caso de un espacio muestral finito, cualquier función $X : \Omega \mapsto \mathbb{R}$ es una variable aleatoria.

PROPOSICIÓN 6.3. *Sea* X *una variable aleatoria real, entonces* $[X \in B] \in \Im$ *para cualquier boreliano* B *de* $\mathbb{R}$.

Demostración

Sea $H = \{B \subset \mathbb{R} : B \text{ es boreliano y } [X \in B] \in \Im\}$, entonces H es una σ-álgebra de subconjuntos de $\mathbb{R}$ que contiene a los intervalos de la forma $(-\infty, x]$, por lo tanto contiene a todos los borelianos de $\mathbb{R}$, ya que, de acuerdo con la definición 5.40, la σ-álgebra de los conjuntos borelianos en $\mathbb{R}$ está generada por los intervalos de ese tipo.

■

PROPOSICIÓN 6.4. *Sean* X *y* Y *dos variables aleatorias reales, entonces* $[(X,Y) \in B] \in \Im$ *para cualquier boreliano* B *de* $\mathbb{R}^2$.

Demostración

Sea $H = \{B \subset \mathbb{R}^2 : B \text{ es boreliano y } [(X,Y) \in B] \in \Im\}$, entonces H es una σ-álgebra de subconjuntos de $\mathbb{R}^2$ que contiene a los rectángulos de la forma $(-\infty, x] \times (-\infty, y]$, por lo tanto contiene a todos los borelianos de $\mathbb{R}^2$, ya que, de acuerdo con la definición 5.55, la σ-álgebra de los conjuntos borelianos en $\mathbb{R}^2$ está generada por los rectángulos de ese tipo.

■

PROPOSICIÓN 6.5. *Sea* X *una variable aleatoria real y* $f : \mathbb{R} \mapsto \mathbb{R}$ *una función boreliana, entonces* $f(X)$ *es una variable aleatoria real.*

Demostración

Sea B un boreliano de $\mathbb{R}$, entonces $f^{-1}(B)$ es boreliano, así que:

$[f(X) \in B] = [X \in f^{-1}(B)] \in \Im$

■

PROPOSICIÓN 6.6. *Sean* X *y* Y *dos variables aleatorias reales y* $f : \mathbb{R}^2 \mapsto \mathbb{R}$ *una función boreliana, entonces* $f(X,Y)$ *es una variable aleatoria real.*

Demostración

Sea B un boreliano de $\mathbb{R}$, entonces $f^{-1}(B)$ es boreliano de $\mathbb{R}^2$, así que:

$[f(X,Y) \in B] = [(X,Y) \in f^{-1}(B)] \in \Im$

■

Corolario 6.7. *Sean X y Y dos variables aleatorias reales y $a, b, \alpha \in \mathbb{R}$ con $\alpha \geq 0$, entonces $aX + b$, $|X|^{\alpha}$, $X + Y$, XY, $\text{máx}(X,Y)$ y $\text{mín}(X,Y)$ también son variables aleatorias reales.*

Proposición 6.8. *Sea $\mathcal{B}$ la σ-álgebra de Borel en $\mathbb{R}$ y X una variable aleatoria real. Entonces la función $\mu_X : \mathcal{B} \mapsto \mathbb{R}$ definida por $\mu_X(B) = P[X \in B]$ es una medida de probabilidad.*

Demostración

Obviamente $\mu_X(\mathbb{R}) = 1$ y $\mu_X(B) \geq 0$ para cualquier $B \in \mathcal{B}$.

Sea $B_1, B_2, \ldots$ una colección finita o infinita numerable de conjuntos borelianos ajenos por parejas, entonces los eventos $P[X \in B_1], P[X \in B_1], \ldots$ son mutuamente excluyentes, así que :

$\mu_X(\cup_k B_k) = P[X \in \cup_k B_k] = P(\cup_k [X \in B_k])$

$= \sum_k P[X \in B_k] = \sum_k \mu_X(B_k)$

Por lo tanto, μ_X es σ-aditiva.

■

El uso explícito del concepto de variable aleatoria real, como fue definido arriba, es relativamente reciente, se dio únicamente a partir de fines del siglo pasado. Sin embargo, implícitamente se han estudiado variables aleatorias casi desde el inicio del Cálculo de Probabilidades como disciplina matemática. Para ser más precisos, ya el trabajo de Christiaan Huygens del año 1657 contiene problemas relativos a variables aleatorias. A medida que se desarrolló la Teoría de la Probabilidad el concepto de variable aleatoria, implícito en muchos problemas, se fue convirtiendo en uno de los conceptos fundamentales.

6.2. Funciones de distribución

Decir que una variable aleatoria puede tomar como valor uno de un conjunto de posibles valores no nos da mucha información pues, como puede verse en los ejemplos, diferentes variables aleatorias pueden admitir el mismo conjunto de posibles valores.

Realmente lo que nos interesa de una variable aleatoria es el valor que va a tomar en un experimento y no únicamente el conjunto de sus posibles valores; sin embargo, esto no podemos conocerlo antes de realizar el experimento. Es por esto que lo más importante de una variable aleatoria es la información sobre las probabilidades con que puede tomar sus diferentes posibles valores. Esta información es la que nos da la función de distribución, concepto que se define a continuación.

DEFINICIÓN 6.9 (**Función de distribución**). *Si X es una variable aleatoria real, la función $F_X : \mathbb{R} \mapsto \mathbb{R}$, definida por $F_X(x) = P[X \leq x]$, es llamada la función de distribución de X.*

EJEMPLO 6.10. *Un experimento aleatorio consiste en seleccionar al azar una bola de una urna, la cual contiene N bolas numeradas del 1 al N, y X es el número de la bola seleccionada, entonces:*

$$F_X(x) = \begin{cases} 0 & si\ x < 1 \\ \frac{1}{N}[[x]] & si\ x \in [1, N) \\ 1 & si\ x \geq N \end{cases}$$

en donde $[[x]]$ es la parte entera de x.

EJEMPLO 6.11. *Un experimento aleatorio consiste en elegir al azar un número del intervalo $(0, 1)$ y X es el número que se obtiene, entonces:*

$$F_X(x) = \begin{cases} 0 & si\ x < 0 \\ x & si\ x \in [0, 1) \\ 1 & si\ x \geq 1 \end{cases}$$

EJEMPLO 6.12. *Un experimento aleatorio consiste en lanzar una moneda indefinidamente y X es el número de caras que se obtienen antes de la primera cruz, entonces:*

$$F_X(x) = \begin{cases} 0 & si\ x < 0 \\ \sum_{k=0}^{[[x]]} \frac{1}{2^{k+1}} & si\ x \geq 0 \end{cases} = \begin{cases} 0 & si\ x < 0 \\ 1 - \frac{1}{2^{[[x]]+1}} & si\ x \geq 0 \end{cases}$$

en donde $[[x]]$ es la parte entera de x.

PROPOSICIÓN 6.13. *Sea X una variable aleatoria real y F_X su función de distribución, entonces:*

(i) *F_X es una función monótona no decreciente y continua por la derecha.*
(ii) $\lim_{x \rightsquigarrow \infty} F_X(x) = 1$
(iii) $\lim_{x \rightsquigarrow -\infty} F_X(x) = 0$

Demostración

i. Sea (x_n) una sucesión monótona decreciente tal que $\lim_{n \rightsquigarrow \infty} x_n = x$, entonces:

$F_X(x+) = \text{lím}_{n\rightsquigarrow\infty} F_X(x_n) = \text{lím}_{n\rightsquigarrow\infty} P[X \leq x_n] = P\left[\bigcap_{n=1}^{\infty}[X \leq x_n]\right]$

$= P[X \leq x] = F_X(x)$

ii. Sea (x_n) una sucesión monótona creciente tal que $\text{lím}_{n\rightsquigarrow\infty} x_n = \infty$, entonces:

$\text{lím}_{n\rightsquigarrow\infty} F_X(x_n) = \text{lím}_{n\rightsquigarrow\infty} P[X \leq x_n] = P\left[\bigcup_{n=1}^{\infty}[X \leq x_n]\right] = P[\Omega] = 1$

iii. Sea (x_n) una sucesión monótona decreciente tal que $\text{lím}_{n\rightsquigarrow\infty} x_n = -\infty$, entonces:

$\text{lím}_{n\rightsquigarrow\infty} F_X(x_n) = \text{lím}_{n\rightsquigarrow\infty} P[X \leq x_n] = P\left[\bigcap_{n=1}^{\infty}[X \leq x_n]\right] = P[\emptyset] = 0$

■

Definición 6.14 (**Función de distribución**). *Se dice que una función $F : \mathbb{R} \mapsto \mathbb{R}$ es una función de distribución si satisface las propiedades i, ii y iii de la última proposición.*

Como se ilustra en los ejemplos, la función de distribución de una variable aleatoria real no siempre es una función continua. La continuidad o discontinuidad de una función de distribución en un punto x depende del valor de $P[X = x]$.

Proposición 6.15. *Sea X una variable aleatoria real y F_X su función de distribución, entonces, para cualquier $x \in \mathbb{R}$, se tiene $F_X(x-) = P[X < x]$.*

Demostración

Sea $x \in \mathbb{R}$ y (x_n) una sucesión monótona creciente tal que $\text{lím}_{n\rightsquigarrow\infty} x_n = x$, entonces:

$F_X(x-) = \text{lím}_{n\rightsquigarrow\infty} F_X(x_n) = \text{lím}_{n\rightsquigarrow\infty} P[X \leq x_n] = P\left[\bigcup_{n=1}^{\infty}[X \leq x_n]\right] = P[X < x]$

■

Corolario 6.16. *Sea X una variable aleatoria real y F_X su función de distribución, entonces:*

$F_X(x) - F_X(x-) = P[X = x]$

Demostración

$F_X(x) - F_X(x-) = P[X \leq x] - P[X < x] = P[X = x]$

■

6.3. Clasificación de variables aleatorias

Lema 6.17. *Si $F : \mathbb{R} \mapsto \mathbb{R}$ es una función monótona no decreciente, entonces el conjunto de puntos en los cuales F es discontinua, es finito o infinito numerable.*

Demostración

Sea D el conjunto de puntos en donde F es discontinua, entonces:

$D = \{x \in \mathbb{R} : F(x+) - F(x-) > 0\}$

Fijemos $m \in \mathbb{N}$ y definamos $A = F(m) - F(-m)$. Entonces, para cada $n \in \mathbb{N}$, el conjunto

$D_n^m = \left\{x \in (-m, m) : F(x+) - F(x-) > \frac{A}{n}\right\}$

tiene a lo más $n-1$ elementos pues, si hubiera n números reales distintos, $x_1, \ldots, x_n$, pertenecientes a D_n^m, se tendría:

$A = F(m) - F(-m) \geq \sum_{k=1}^{n} [F(x_k+) - F(x_k-)] > \frac{nA}{n} = A$

Por lo tanto, el conjunto $D^m = \{x \in (-m, m) : F(x+) - F(x-) > 0\} = \bigcup_{n=1}^{\infty} D_n^m$

es finito o infinito numerable.

Así que, $D = \bigcup_{m=1}^{\infty} D^m$ también es finito o infinito numerable.

$\blacksquare$

Corolario 6.18. *El conjunto $\{x \in \mathbb{R} : P[X = x] > 0\}$ es finito o infinito numerable.*

Demostración

De acuerdo con la proposición 6.16, la función de distribución de una variable aleatoria real X es discontinua en un punto x si y sólo si $P[X = x] > 0$, de manera que el resultado se sigue inmediatamente del lema 6.17.

$\blacksquare$

Sea X una variable aleatoria y definamos:

$D = \{x \in \mathbb{R} : P[X = x] > 0\}$

$C = \{x \in \mathbb{R} : P[X = x] = 0\}$

$p = P[X \in D]$

Entonces, de acuerdo con la última proposición, D es numerable.

Las variables aleatorias se clasifican de acuerdo al valor de p. Si $p = 0$, diremos que la variable aleatoria es continua, si $p = 1$ diremos que es discreta. Cuando $0 < p < 1$, se tiene $P[X \in D] > 0$ y $P[X \in C] > 0$, de manera que se puede decir que, en ese caso, la variable aleatoria tiene una parte discreta (la que corresponde al conjunto D) y una parte continua (la que corresponde al conjunto C).

Definición 6.19 (**Variable aleatoria continua**). *Se dice que una variable aleatoria es continua si su función de distribución es continua.*

Obsérvese que una variable aleatoria X es continua si y sólo si $P[X = x] = 0$ para cualquier $x \in \mathbb{R}$.

Entre las funciones continuas $F : \mathbb{R} \mapsto \mathbb{R}$, son de particular importancia aquellas de la forma $F(x) = \int_{-\infty}^{x} f(y)dy$ para alguna función integrable f. Este tipo de funciones continuas son llamadas absolutamente continuas.

Definición 6.20 (**Variable aleatoria absolutamente continua**). *Se dice que la función de distribución F_X de la variable aleatoria X es absolutamente continua si existe una función boreliana no negativa $f_X : \mathbb{R} \mapsto \mathbb{R}$, integrable, tal que:*

$F_X(x) = \int_{-\infty}^{x} f_X(y)dy$ *para cualquier* $x \in \mathbb{R}$.

En este caso se dice también que la variable aleatoria X es absolutamente continua y la función f_X es llamada una función de densidad de X.

Cuando existe una función de densidad de una variable aleatoria continua X, ésta no es única. En efecto, dada una de ellas, se puede, por ejemplo, modificar su valor en un número finito de puntos y la nueva función que se obtiene sigue siendo una función de densidad de X.

Dado $x \in \mathbb{R}$, definamos $A = (-\infty, x]$. La propiedad que caracteriza a una función de densidad de X se puede escribir entonces de la siguiente manera:

$$P[X \in A] = \int_A f_X(y)dy$$

Cabe entonces preguntarse si esta propiedad puede extenderse a una familia más grande de conjuntos A.

La integral de Riemann presenta limitaciones para el tratamiento de las probabilidades $P[X \in A]$ como integrales $\int_A f_X(y)dy$ pues la integral puede no estar definida aún cuando la probabilidad $P[X \in A]$ sí lo esté. Por ejemplo, si f_X es positiva y constante en un intervalo (a, b) y A es el conjunto de número racionales en ese intervalo, entonces $P[X \in A] = 0$, pero la integral $\int_A f_X(y)dy = \int_a^b I_A(y)f_X(y)dy$ no está definida como integral de Riemann pues el producto I_Af_X es una función discontinua en todos los puntos del intervalo (a, b).

Lo más adecuado para tratar las probabilidades $P[X \in A]$ como integrales $\int_A f_X(y)dy$ es utilizar la definición de Lebesgue de la integral, tema que se desarrolla en el Apéndice.

PROPOSICIÓN 6.21. *Sea X una variable aleatoria absolutamente continua y f_X una función de densidad de X. Entonces $P[X \in B] = \int_B f_X(x)dx$ para cualquier conjunto boreliano B.*

Demostración

Sea $\mathcal{B}$ la σ-álgebra de Borel en $\mathbb{R}$ y definamos la función $\mu : \mathcal{B} \mapsto \mathbb{R}$ de la siguiente manera:

$\mu(B) = \int_B f_X(x)dx$

Obviamente $\mu(\mathbb{R}) = 1$, $\mu_X(B) \geq 0$ para cualquier $B \in \mathcal{B}$ y μ es finitamente aditiva.

Sea $B_1, B_2, \ldots$ una sucesión monótona no decreciente de conjuntos borelianos, entonces la sucesión de funciones $I_{B_n} f_X$ es monótona no decreciente, así que por el teorema de la convergencia monótona, se tiene:

$\mu(\cup_{n=1}^{\infty} B_n) = \int_{\cup_{n=1}^{\infty} B_n} f_X(x)dx = \int_{\mathbb{R}} I_{\cup_{n=1}^{\infty} B_n}(x) f_X(x)dx$

$= \int_{\mathbb{R}} \lim_{n \rightsquigarrow \infty} I_{B_n}(x) f_X(x)dx = \lim_{n \rightsquigarrow \infty} \int_{\mathbb{R}} I_{B_n}(x) f_X(x)dx$

$= \lim_{n \rightsquigarrow \infty} \int_{B_n} f_X(x)dx = \lim_{n \rightsquigarrow \infty} \mu(B_n)$

Así que μ es σ-aditiva y, por lo tanto, es una medida de probabilidad.

Por otra parte, por la proposición 6.8, la función $\mu_X : \mathcal{B} \mapsto \mathbb{R}$, definida por $\mu_X(B) = P[X \in B]$ es una medida de probabilidad.

Además:

$\mu_X((-\infty, x]) = P[X \leq x] = \int_{-\infty}^{x} f_X(y)dy = \mu((-\infty, x])$

para cualquier $x \in \mathbb{R}$

Así que, por la proposición 5.39, $\mu_X(B) = \mu(B)$ para cualquier $B \in \mathcal{B}$, lo cual prueba el resultado.

■

La continuidad absoluta de una variable aleatoria puede establecerse aplicando el siguiente resultado, el cual es un corolario del Teorema Fundamental del Cálculo:

LEMA 6.22. *Sea $F : \mathbb{R} \mapsto \mathbb{R}$ una función de distribución continua. Sean $a = \text{ínf}\{x \in \mathbb{R} : F(x) > 0\}$ y $b = \sup\{x \in \mathbb{R} : F(x) < 1\}$, y supongamos que la derivada F' de F existe y es continua en el intervalo (a, b) de tal manera que cuando a (resp. b) es finito, F' se puede extender continuamente a a (resp. b), entonces la función $f : \mathbb{R} \mapsto \mathbb{R}$ definida por:*

$$f(x) = \begin{cases} F'(x) & si\ x \in (a, b) \\ 0 & en\ otro\ caso \end{cases}$$

es una función de densidad de F.

PROPOSICIÓN 6.23. *Sea X una variable aleatoria absolutamente continua y f_X una función de densidad de X, entonces:*

(i) $f_X(x) \geq 0$ *para cualquier* $x \in \mathbb{R}$.
(ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$

Demostración

La primera propiedad es parte de la definición de función de densidad. Para la segunda, se tiene:

$\int_{-\infty}^{\infty} f_X(x)dx = \text{lím}_{y \rightsquigarrow \infty} \int_{-\infty}^{y} f_X(x)dx = \text{lím}_{y \rightsquigarrow \infty} F_X(y) = 1$

■

DEFINICIÓN 6.24 (**Función de densidad**). *Se dice que una función boreliana $f : \mathbb{R} \mapsto \mathbb{R}$ es una función de densidad si satisface las propiedades i y ii de la última proposición.*

EJEMPLO 6.25. *Las siguientes son funciones de densidad:*

$$f(x) = \begin{cases} (n+1)x^n & si\ x \in (0, 1) \\ 0 & en\ otro\ caso \end{cases}$$

en donde n es un entero no negativo.

$$f(x) = \begin{cases} (n-1)x^{-n} & si\ x \in (1, \infty) \\ 0 & en\ otro\ caso \end{cases}$$

en donde $n \in \{2, 3, \ldots\}$.

$$f(x) = \begin{cases} \frac{1}{3} & si\ x \in (0, 1) \\ \frac{1}{6} & si\ x \in (2, 3) \\ \frac{1}{2} & si\ x \in (4, 5) \\ 0 & en\ otro\ caso \end{cases}$$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & si\ x > 0 \\ 0 & en\ otro\ caso \end{cases}$$

en donde $\lambda > 0$*.*

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ *para cualquier* $x \in \mathbb{R}$*.*

DEFINICIÓN 6.26 (**Variable aleatoria discreta**). *Se dice que una variable aleatoria real* X *es discreta si existe una colección finita o infinita numerable de números reales,* $x_1, x_2, \ldots$*, tales que* $P[X = x_k] > 0$ *para cualquier* x_k *y* $\sum_k P[X = x_k] = 1$*. En este caso, diremos que el conjunto* $V_X = \{x_1, x_2, \ldots\}$ *es el conjunto de posibles valores de* X*.*

EJEMPLO 6.27. *Si* $A_1, A_2, \ldots$ *es una colección finita o numerable de eventos mutuamente excluyentes y* $a_1, a_2, \ldots$ *son números reales arbitrarios, entonces la función* $X = \sum_k a_k I_{A_k}$ *es una variable aleatoria discreta.*

En particular, Si A *es un evento, entonces* $X = I_A$ *es una variable aleatoria discreta.*

▲

El ejemplo anterior da básicamente la forma general de una variable aleatoria discreta X pues si $V_X = \{x_1, x_2, \ldots\}$ es su conjunto de posibles valores y definimos $Y = \sum_k x_k I_{[X=x_k]}$, entonces $P[X = Y] = 1$.

PROPOSICIÓN 6.28. *Sea* X *una variable aleatoria discreta, entonces la probabilidad de que* X *tome un valor finito es igual a* 1*.*

Demostración

Sea $V_X = \{x_1, x_2, \ldots\}$ el conjunto de los posibles valores de X, entonces:

$P[X \in V_X^c] = 1 - P[X \in V_X] = 1 - \sum_k P[X = x_k] = 0$

Por lo tanto:

$P[X < \infty] = P[X < \infty, X \in V_X] + P[X < \infty, X \in V_X^c]$

$= P[X \in V_X] = \sum_k P[X = x_k] = 1$

■

DEFINICIÓN 6.29 (**Función de densidad de una variable aleatoria discreta**). *Sea* X *una variable aleatoria discreta. La función* $f_X : \mathbb{R} \mapsto \mathbb{R}$ *definida por* $f(x) = P[X = x]$ *es llamada la función de densidad (o función de probabilidad) de* X*.*

La definición 6.26 implica inmediatamente el siguiente resultado:

Proposición 6.30. *Sea X una variable aleatoria discreta y f_X su función de densidad, entonces:*

(i) $0 \leq f_X(x) \leq 1$
(ii) *Existe un conjunto finito o infinito numerable D tal que $f_X(x) > 0$ para cualquier $x \in D$ y $\sum_{x \in D} f_X(x) = 1$.*

Definición 6.31 (**Función de densidad discreta**). *Se dice que una función $f : \mathbb{R} \mapsto \mathbb{R}$ es una función de densidad discreta si satisface las propiedades i y ii de la última proposición.*

Ejemplo 6.32. *Las siguientes son funciones de densidad discretas:*

$$f(x) = \begin{cases} cr^x & \text{si } x \in \{0, \ldots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde $r > 0$, $N \in \mathbb{N}$ *y* $c = \frac{1}{\sum_{x=0}^{N} r^x} = \begin{cases} \frac{1}{N+1} & \text{si } r = 1 \\ \frac{1-r}{1-r^{N+1}} & \text{si } r \neq 1 \end{cases}$

$$f(x) = \begin{cases} cr^x & \text{si } x \in \{0, 1, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde $0 < r < 1$ *y* $c = \frac{1}{\sum_{x=0}^{\infty} r^x} = 1 - r$.

$$f(x) = \begin{cases} cr^{\sqrt{x}} & \text{si } x \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$$

en donde $0 < r < 1$ *y* $c = \frac{1}{\sum_{x=0}^{\infty} r^{\sqrt{x}}}$.

$$f(x) = \begin{cases} \frac{c}{x^2} & \text{si } x \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$$

en donde $k \in \{2, 3, \ldots\}$ *y* $c = \frac{1}{\sum_{x=1}^{\infty} \frac{1}{x^2}} = \frac{6}{\pi^2}$.

$$f(x) = \begin{cases} \frac{1}{x(x+1)} & \text{si } x \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{2^k} & \text{si } x = 2^k,\ k \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{k(k+1)2^k} & \text{si } x = 2^k,\ k \in \mathbb{N} \\ 1 - \sum_{k=1}^{\infty} \frac{1}{k(k+1)2^k} & \text{si } x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$f(x)=\begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & si\ x\in\{0,1,\ldots\} \\ 0 & en\ otro\ caso \end{cases}$$

en donde $\lambda > 0$.

▲

Una variable aleatoria real no necesariamente cae en una de las dos categorías definidas arriba. En general, dada una variable aleatoria X, existe un conjunto finito o infinito numerable D tal que $P[X=x]>0$ para cualquier $x\in D$, $P[X=x]=0$ para cualquier $x\notin D$ y $\sum_{x\in D}P[X=x]\leq 1$. Cuando $\sum_{x\in D}P[X=x]=1$, la variable aleatoria es discreta, mientras que cuando $D=\emptyset$, la variable aleatoria es continua. Pero, si $D\neq\emptyset$ y $\sum_{x\in D}P[X=x]<1$, la variable aleatoria no es ni discreta ni continua.

Ejemplo 6.33. *Sea X una variable aleatoria con función de distribución dada por:*

$$F_X(x)=\begin{cases} 0 & si\ x<0 \\ 1-e^{-x} & si\ 0\leq x<1 \\ 1 & si\ x\geq 1 \end{cases}$$

entonces $P[X=1]=e^{-1}$ y $P[X=x]=0$ para cualquier $x\neq 1$, de manera que X no es ni discreta ni continua.

▲

Como ya lo mencionamos, una variable aleatoria real no necesariamente es discreta o continua, pero, lo que sí se puede afirmar es que toda función de distribución tiene una parte de saltos y una parte continua, lo cual hace que se pueda decir que toda variable aleatoria está compuesta por una parte discreta y una parte continua. Este resultado se prueba a continuación:

Proposición 6.34. *Sea f una función de densidad discreta y definamos la función $F:\mathbb{R}\mapsto\mathbb{R}$ de la siguiente manera:*

$$F(x)=\sum\nolimits_{\{y\in\mathbb{R}:y\leq x,f(y)>0\}}f(y)$$

Entonces F es una función de distribución y $F(x-)=\sum_{\{y\in\mathbb{R}:y<x,f(y)>0\}}f(y)$.

Demostración

F está bien definida ya que la serie $\sum_{\{y\in\mathbb{R}:f(y)>0\}}f(y)$ es convergente. Además, obviamente es una función monótona no decreciente.

Sea $D=\{y\in\mathbb{R}:f(y)>0\}$ y, para cada $x\in\mathbb{R}$, definamos $D_x=\{y\in D:y\leq x\}$.

Si x_n es una sucesión creciente que tiende a ∞, se tiene $\cup_{n=1}^{\infty} D_{x_n} = D$, así que:

$\lim_{n \rightsquigarrow \infty} F(x_n) = \lim_{n \rightsquigarrow \infty} \sum_{y \in D_{x_n}} f(y) = \sum_{y \in D} f(y) = 1$

Si x_n es una sucesión decreciente que tiende a $-\infty$, se tiene $\cap_{n=1}^{\infty} D_{x_n} = \varnothing$, así que:

$\lim_{n \rightsquigarrow \infty} F(x_n) = \lim_{n \rightsquigarrow \infty} \sum_{\{y \in D: y \leq x_n\}} f(y) = 1 - \lim_{n \rightsquigarrow \infty} \sum_{\{y \in D: y > x_n\}} f(y)$

$= 1 - \lim_{n \rightsquigarrow \infty} \sum_{y \in D - D_{x_n}} f(y) = 1 - \sum_{y \in D} f(y) = 0$

Si x_n es una sucesión decreciente que tiende a x, se tiene $\cap_{n=1}^{\infty} D_{x_n} = D_x$, así que:

$\lim_{n \rightsquigarrow \infty} F(x_n) = \lim_{n \rightsquigarrow \infty} \sum_{\{y \in D: y \leq x_n\}} f(y) = 1 - \lim_{n \rightsquigarrow \infty} \sum_{\{y \in D: y > x_n\}} f(y)$

$= 1 - \lim_{n \rightsquigarrow \infty} \sum_{y \in D - D_{x_n}} f(y) = 1 - \sum_{y \in D - D_x} f(y) = 1 - \sum_{\{y \in D: y > x\}} f(y) = F(x)$

Por lo tanto, F es continua por la derecha.

Si x_n es una sucesión creciente que tiende a x, se tiene $\cup_{n=1}^{\infty} D_{x_n} = \{y \in D : y < x\}$, así que:

$\lim_{n \rightsquigarrow \infty} F(x_n) = \lim_{n \rightsquigarrow \infty} \sum_{y \in D_{x_n}} f(y) = \sum_{\{y \in D: y < x\}} f(y) = \sum_{\{y \in \mathbb{R}: y < x, f(y) > 0\}} f(y)$

■

Definición 6.35 (**Función de distribución discreta**). *Se dice que una función $F : \mathbb{R} \mapsto \mathbb{R}$ es una función de distribución discreta si existe una función de densidad discreta f tal que:*

$$F(x) = \sum_{\{y \in \mathbb{R}: y \leq x, f(y) > 0\}} f(x)$$

para cualquier $x \in \mathbb{R}$.

Proposición 6.36. *Sea X una variable aleatoria tal que:*

$$0 < \sum_{\{x \in \mathbb{R}: P[X=x] > 0\}} P[X = x] < 1$$

entonces su función de distribución F_X puede expresarse en la forma:

$$F_X = \alpha F_X^d + (1 - \alpha) F_X^c$$

en donde $\alpha \in (0, 1)$, F_X^d es una función de distribución discreta y F_X^c es una función de distribución continua. Además, esta descomposición es única.

Demostración

Sea $\alpha = \sum_{\{x \in \mathbb{R}: P[X=x] > 0\}} P[X = x]$, entonces la función $f : \mathbb{R} \mapsto \mathbb{R}$ definida por:

$f(x) = \frac{1}{\alpha} P[X = x]$

es una función de densidad discreta. Por lo tanto, la función $F_X^d : \mathbb{R} \mapsto \mathbb{R}$ definida por:

$F_X^d(x) = \sum_{\{y \in \mathbb{R}: y \leq x, f(y) > 0\}} f(y)$

es una función de distribución discreta.

Definamos F_X^c de la siguiente manera:

$F_X^c = \frac{1}{1-\alpha} \left(F_X - \alpha F_X^d \right)$

Se tiene:

$\text{lím}_{x \rightsquigarrow \infty} F_X^c(x) = \frac{1}{1-\alpha} \left[\text{lím}_{x \rightsquigarrow \infty} F_X(x) - \alpha \, \text{lím}_{x \rightsquigarrow \infty} F_X^d(x) \right]$

$= \frac{1}{1-\alpha} (1 - \alpha) = 1$

$\text{lím}_{x \rightsquigarrow -\infty} F_X^c(x) = \frac{1}{1-\alpha} \left[\text{lím}_{x \rightsquigarrow -\infty} F_X(x) - \alpha \, \text{lím}_{x \rightsquigarrow -\infty} F_X^d(x) \right] = 0$

Si $x < y$ entonces:

$\left[F_X(y) - \alpha F_X^d(y) \right] - \left[F_X(x) - \alpha F_X^d(x) \right]$

$= [F_X(y) - F_X(x)] - \alpha \left[F_X^d(y) - F_X^d(x) \right]$

$= [F_X(y) - F_X(x)] - \alpha \sum_{\{y \in \mathbb{R}: x < z \leq y, f(z) > 0\}} f(z)$

$= P[x < X \leq y] - \sum_{\{y \in \mathbb{R}: x < z \leq y, P[X = z] > 0\}} P[X = z] \geq 0$

Así que $F_X^c(x) \leq F_X^c(y)$.

$\left[F_X(x+) - \alpha F_X^d(x+) \right] - \left[F_X(x) - \alpha F_X^d(x) \right]$

$= [F_X(x+) - F_X(x)] - \alpha \left[F_X^d(x+) - F_X^d(x) \right] = 0$

Así que F_X^c es continua por la derecha.

$\left[F_X(x) - \alpha F_X^d(x) \right] - \left[F_X(x-) - \alpha F_X^d(x-) \right]$

$= [F_X(x) - F_X(x-)] - \alpha \left[F_X^d(x) - F_X^d(x-) \right]$

$= P[X = x] - P[X = x] = 0$

Así que F_X^c es continua por la izquierda.

Por lo tanto, F_X^c es una función de distribución continua.

Para la unicidad, sea $F_X = \beta F^d + (1-\beta) F^c$ otra descomposición. Se tiene entonces:

$\alpha F_X^d + (1-\alpha) F_X^c = \beta F^d + (1-\beta) F^c$

$\alpha F_X^d - \beta F^d = (1-\beta) F^c - (1-\alpha) F_X^c$

$\alpha f(x) - \beta g(x) = 0$

$f(x) = \frac{\beta}{\alpha} g(x)$

Así que $\frac{\beta}{\alpha} = 1$ y $f = g$. Por lo tanto $F_X^d = F^d$ y $F_X^c = F^c$. ■

COROLARIO 6.37. *Sea X cualquier variable aleatoria, entonces su función de distribución F_X puede expresarse en la forma:*

$$F_X = \alpha F_X^d + (1-\alpha) F_X^c$$

en donde $\alpha \in [0,1]$, F_X^d es una función de distribución discreta y F_X^c es una función de distribución continua.

Demostración

Si $0 < \sum_{\{x \in \mathbb{R}: P[X=x]>0\}} P[X=x] < 1$, consideramos la descomposición de la proposición anterior.

Si $\sum_{\{x \in \mathbb{R}: P[X=x]>0\}} P[X=x] = 1$, tomamos $\alpha = 1$, $F_X^d = F_X$ y $F_X^c = 0$.

Si $\sum_{\{x \in \mathbb{R}: P[X=x]>0\}} P[X=x] = 0$, tomamos $\alpha = 0$, $F_X^d = 0$ y $F_X^c = F_X$.

■

6.4. Independencia de variables aleatorias

DEFINICIÓN 6.38 (**Independencia de variables aleatorias**). *Se dice que n variables aleatorias, $X_1, \ldots, X_n$, son independientes si para cualquier colección de subconjuntos borelianos de números reales, $A_1, \ldots, A_n$, se tiene:*

$$P[X_1 \in A_1, \ldots, X_n \in A_n] = P[X_1 \in A_1] \cdots P[X_n \in A_n]$$

Obsérvese que la condición de independencia de $X_1, \ldots, X_n$ equivale a pedir que, para cualquier colección de subconjuntos de números reales, $A_1, \ldots, A_n$, los eventos $[X_1 \in A_1]$, ..., $[X_n \in A_n]$ son independientes.

Proposición 6.39. *Sean $X_1, \ldots, X_n$ n variables aleatorias independientes, entonces cualquier subcolección de ellas, $X_{i_1}, \ldots, X_{i_k}$, es también una familia de variables aleatorias independientes.*

Demostración

Sean $A_{i_1}, \ldots, A_{i_k}$ subconjuntos borelianos de $\mathbb{R}$ y, para $j \in \{1, \ldots, n\} - \{i_1, \ldots, i_k\}$, definamos $A_j = \mathbb{R}$, entonces:

$$P\left[X_{i_1} \in A_{i_1}, \ldots, X_{i_k} \in A_{i_k}\right] = P\left[X_1 \in A_1, \ldots, X_n \in A_n\right]$$

$$= P\left[X_1 \in A_1\right] \cdots P\left[X_n \in A_n\right] = P\left[X_{i_1} \in A_{i_1}\right] \cdots P\left[X_{i_k} \in A_{i_k}\right]$$

■

Proposición 6.40. *Sean $X_1, \ldots, X_n$ n variables aleatorias independientes y $f_1, \ldots, f_n$ n funciones borelianas de $\mathbb{R}$ en $\mathbb{R}$. Entonces las variables aleatorias $f_1(X_1), \ldots, f_n(X_n)$ son independientes.*

Demostración

Sean $A_1, \ldots, A_n$ subconjuntos borelianos de $\mathbb{R}$, entonces $f_1^{-1}(A_1), \ldots, f_n^{-1}(A_n)$ son también subconjuntos borelianos de $\mathbb{R}$, así que:

$$P\left[f_1(X_1) \in A_1, \ldots, f_n(X_n) \in A_n\right] = P\left[X_1 \in f_1^{-1}(A_1), \ldots, X_n \in f_n^{-1}(A_n)\right]$$

$$= P\left[X_1 \in f_1^{-1}(A_1)\right] \cdots P\left[X_n \in f_n^{-1}(A_n)\right] = P\left[f_1(X_1) \in A_1\right] \cdots P\left[f_n(X_n) \in A_n\right]$$

■

Proposición 6.41. *Sean $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m}$ $n+m$ variables aleatorias independientes y $f : \mathbb{R}^n \mapsto \mathbb{R}$ y $g : \mathbb{R}^m \mapsto \mathbb{R}$ dos funciones borelianas. Entonces, las variables aleatorias $f(X_1, \ldots, X_n)$ y $g(X_{n+1}, \ldots, X_{n+m})$ son independientes.*

Demostración

Sea $\mathcal{G}$ la familia de subconjuntos $B \subset \mathbb{R}^m$ tales que:

$$P\left[(X_1, \ldots, X_n) \in I, (X_{n+1}, \ldots, X_{n+m}) \in B\right]$$

$$= P\left[(X_1, \ldots, X_n) \in I\right] P\left[(X_{n+1}, \ldots, X_{n+m}) \in B\right]$$

para cualquier intervalo I de $\mathbb{R}^n$.

$\mathcal{G}$ es entonces un d-sistema que contiene a los rectángulos de $\mathbb{R}^m$, así que, por el teorema 5.38, $\mathcal{G}$ contiene a todos los borelianos de $\mathbb{R}^m$.

Sea $\mathcal{H}$ la familia de subconjuntos $A \subset \mathbb{R}^n$ tales que:

$P[(X_1,\ldots,X_n)\in A,(X_{n+1},\ldots,X_{n+m})\in B]$

$= P[(X_1,\ldots,X_n)\in A]\,P[(X_{n+1},\ldots,X_{n+m})\in B]$

para cualquier boreliano B de $\mathbb{R}^m$.

$\mathcal{H}$ es entonces un d-sistema que contiene a los rectángulos de $\mathbb{R}^n$, así que, por el teorema 5.38, $\mathcal{H}$ contiene a todos los borelianos de $\mathbb{R}^n$.

Por lo tanto:

$P[(X_1,\ldots,X_n)\in A,(X_{n+1},\ldots,X_{n+m})\in B]$

$= P[(X_1,\ldots,X_n)\in A]\,P[(X_{n+1},\ldots,X_{n+m})\in B]$

para cualquier pareja de borelianos A y B de $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente.

Sean C y D subconjuntos borelianos de $\mathbb{R}$, entonces $f^{-1}(C)$ y $g^{-1}(D)$ son subconjuntos borelianos de $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente. Por lo tanto:

$P[f(X_1,\ldots,X_n)\in C, g(X_{n+1},\ldots,X_{n+m})\in D]$

$= P[(X_1,\ldots,X_n)\in f^{-1}(C),(X_{n+1},\ldots,X_{n+m})\in g^{-1}(D)]$

$= P[(X_1,\ldots,X_n)\in f^{-1}(C)]\,P[(X_{n+1},\ldots,X_{n+m})\in g^{-1}(D)]$

$= P[f(X_1,\ldots,X_n)\in C]\,P[g(X_{n+1},\ldots,X_{n+m})\in D]$

■

Las definiciones y resultados de las siguientes dos secciones se utilizarán más adelante para el estudio de las variables aleatorias.

6.5. Función gama

Proposición 6.42. *La integral $\int_0^\infty t^{\alpha-1}e^{-t}dt$ es finita para cualquier $\alpha>0$.*

Demostración

Obsérvese que $\lim_{t\rightsquigarrow\infty} t^{\alpha+1}e^{-t}=0$, así que, existe $K>0$ tal que $t^{\alpha+1}e^{-t}<1$ para cualquier $t\geq K$. Por lo tanto:

$\int_K^\infty t^{\alpha-1}e^{-t}dt=\int_K^\infty t^{-2}t^{\alpha+1}e^{-t}dt<\int_K^\infty t^{-2}dt=\frac{1}{K}<\infty$

Además:

$\int_0^K t^{\alpha-1}e^{-t}dt < \int_0^K t^{\alpha-1}dt = \frac{1}{\alpha}K^\alpha < \infty$

$\blacksquare$

Definición 6.43 (**Función gama**). *La función* $\Gamma : (0, \infty) \mapsto R$ *definida por* $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$, *cuya gráfica se muestra a continuación, es llamada la función gama.*

Proposición 6.44. $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ *para cualquier* $\alpha > 0$.

Demostración

Integrando por partes se obtiene:

$\Gamma(\alpha+1) = \int_0^\infty t^\alpha e^{-t}dt = \alpha\int_0^\infty t^{\alpha-1}e^{-t}dt = \alpha\Gamma(\alpha)$

$\blacksquare$

Corolario 6.45. *Si* n *es un entero positivo, entonces* $\Gamma(n) = (n-1)!$

Por otra parte:

$\Gamma(\frac{1}{2}) = \int_0^\infty t^{-\frac{1}{2}}e^{-t}dt = 2\int_0^\infty e^{-x^2}dy$

Sea $A = \int_{-\infty}^\infty e^{-x^2}dx$, entonces:

$A^2 = \left[\int_{-\infty}^\infty e^{-x^2}dx\right]\left[\int_{-\infty}^\infty e^{-y^2}dy\right] = \int_{-\infty}^\infty\int_{-\infty}^\infty e^{-x^2}e^{-y^2}dxdy$

Ahora, cambiando a coordenadas polares, se obtiene:

$A^2 = \int_0^\infty\int_0^{2\pi} re^{-r^2}d\theta dr = \pi\int_0^\infty 2re^{-r^2}dr = \pi$

Por lo tanto:

$\int_{-\infty}^\infty e^{-x^2}dx = \sqrt{\pi}$

Así que:

$\Gamma(\frac{1}{2}) = 2\int_0^\infty e^{-x^2}dx = \int_{-\infty}^\infty e^{-x^2}dx = \sqrt{\pi}$

Mediante la función gama podemos expresar de otra forma de definición de los coeficientes binomiales (véase la sección 3.4).

Proposición 6.46. *Sean $x \in \mathbb{R}$ y $n \in \{0, 1, \ldots\}$, entonces:*

$$\binom{z}{m} = \begin{cases} \frac{\Gamma(z+1)}{m!\Gamma(z-m+1)} & si\ m \leq z \\ (-1)^m \frac{\Gamma(m-z)}{m!\Gamma(-z)} & si\ z < 0 \\ (-1)^{[[m-z]]} \frac{\Gamma(z+1)\Gamma(m-z)}{m!} \frac{\langle z\rangle(1-\langle z\rangle)}{\Gamma(1+\langle z\rangle)\Gamma(2-\langle z\rangle)} & si\ 0 \leq z < m \end{cases}$$

en donde $\langle z\rangle$ representa a la parte decimal de z y $[[m-z]]$ a la parte entera de $m-z$.

Demostración

Para $m = 0$, de acuerdo con la definición 3.11, $\binom{z}{m} = 1$ para cualquier $z \in \mathbb{R}$, así que el resultado se cumple trivialmente.

Si $0 < m \leq z$, se tiene:

$\Gamma(z+1) = z(z-1)\cdots(z-m+1)\Gamma(z-m+1)$

Así que:

$\binom{z}{m} = \frac{z(z-1)(z-2)\cdots(z-m+1)}{m!} = \frac{\Gamma(z+1)}{m!\Gamma(z-m+1)}$

Si $z < 0 < m$, entonces $0 < m \leq m-1-z$, así que, de acuerdo con la proposición 3.12 y el resultado previo, se tiene:

$\binom{z}{m} = (-1)^m \binom{m-1-z}{m} = (-1)^m \frac{\Gamma(m-z)}{m!\Gamma(-z)}$

Si $0 \leq z < m$, se tiene:

$\Gamma(z+1) = z(z-1)\cdots(1+\langle z\rangle)\,\Gamma(1+\langle z\rangle)$

$\Gamma(m-z) = (m-z-1)(m-z-2)\cdots(2-\langle z\rangle)\,\Gamma(2-\langle z\rangle)$

$= (-1)^{[[m-z]]-1}\,(\langle z\rangle - 2)\cdots(z-m+1)\Gamma\,(2-\langle z\rangle)$

Así que:

$\binom{z}{m} = \frac{z(z-1)(z-2)\cdots(z-m+1)}{m!} = (-1)^{[[m-z]]} \frac{\Gamma(z+1)\Gamma(m-z)}{m!} \frac{\langle z\rangle(1-\langle z\rangle)}{\Gamma(1+\langle z\rangle)\Gamma(2-\langle z\rangle)}$

■

6.6. Fórmulas de Wallis y de Stirling

Teorema 6.47 (**Fórmula de Wallis**). $\lim_{n \rightsquigarrow \infty} \frac{2\cdot2\cdot4\cdot4\cdot6\cdot6\cdots(2n)2n)}{1\cdot3\cdot3\cdot5\cdot5\cdots(2n-1)(2n-1)(2n+1)} = \frac{\pi}{2}$

Demostración

Para $n \in \{0, 1, \ldots\}$, sea $S_n = \int_0^{\frac{\pi}{2}} sen^n x dx$.

Como, para $x \in (0, \frac{\pi}{2})$, $0 < senx < 1$, la sucesión $\{S_n\}$ es monótona decreciente y $S_n > 0$ para cualquier n.

Por otra parte, integrando por partes, se obtiene, para $n \geq 2$:

$\int sen^n x dx = -\frac{1}{n} sen^{n-1} x \cos x + \frac{n-1}{n} \int sen^{n-2} x dx$

Así que:

$S_n = \int_0^{\frac{\pi}{2}} (senx)^n dx = \frac{n-1}{n} \int sen^{n-2} x dx = \frac{n-1}{n} S_{n-2}$

Por lo tanto, $\frac{S_n}{S_{n+2}} = \frac{n+2}{n+1}$ para cualquier n.

Además, como la sucesión $\{S_n\}$ es monótona decreciente, se tiene $\frac{S_n}{S_{n+1}} \geq 1$ y $\frac{S_n}{S_{n+1}} \leq \frac{S_n}{S_{n+2}}$ para cualquier n. Así que:

$1 \leq \frac{S_n}{S_{n+1}} \leq \frac{S_n}{S_{n+2}} = \frac{n+2}{n+1}$

Por lo tanto, $\lim_{n \rightsquigarrow \infty} \frac{S_n}{S_{n+1}} = 1$

También se tiene, para $n \geq 1$:

$S_{2n} = \frac{2n-1}{2n} S_{2n-2} = \frac{2n-1}{2n} \frac{2n-3}{2n-2} S_{2n-4}$

$= \cdots = \frac{2n-1}{2n} \frac{2n-3}{2n-2} \cdots \frac{1}{2} S_0 = \frac{2n-1}{2n} \frac{2n-3}{2n-2} \cdots \frac{1}{2} \frac{\pi}{2} = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)} \frac{\pi}{2}$

$S_{2n+1} = \frac{2n}{2n+1} S_{2n-1} = \frac{2n}{2n+1} \frac{2n-2}{2n-1} S_{2n-3}$

$= \cdots = \frac{2n}{2n+1} \frac{2n-2}{2n-1} \cdots \frac{2}{3} S_1 = \frac{2n}{2n+1} \frac{2n-2}{2n-1} \cdots \frac{2}{3} = \frac{2 \cdot 4 \cdot 6 \cdots (2n)}{3 \cdot 5 \cdots (2n+1)}$

Así que:

$\frac{S_{2n}}{S_{2n+1}} = \frac{1 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdots (2n-1)(2n-1)(2n+1)}{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \cdots (2n) 2n)} \frac{\pi}{2}$

Por lo tanto:

$\lim_{n \rightsquigarrow \infty} \frac{1 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdots (2n-1)(2n-1)(2n+1)}{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \cdots (2n) 2n)} \frac{\pi}{2} = \lim_{n \rightsquigarrow \infty} \frac{S_{2n}}{S_{2n+1}} = 1$

de lo cual se sigue el resultado.

∎

Corolario 6.48. $\lim_{n\rightsquigarrow\infty} \frac{2^{2n}(n!)^2}{(2n)!\sqrt{n}} = \sqrt{\pi}$

Demostración

$$\frac{2^{2n}(n!)^2}{(2n)!\sqrt{n}} = \frac{2^{2n}(n!)^2}{1\cdot 2\cdot 3\cdots(2n)\sqrt{n}} = \frac{2^{2n}(n!)^2}{1\cdot 3\cdot 5\cdots(2n-1)\cdot 2\cdot 4\cdot 6\cdots(2n)\sqrt{n}}$$

$$= \frac{2^{2n}(n!)^2}{1\cdot 3\cdot 5\cdots(2n-1)\cdot n!2^n\sqrt{n}} = \frac{2^n n!}{1\cdot 3\cdot 5\cdots(2n-1)\sqrt{n}} = \frac{2\cdot 4\cdot 6\cdots(2n)}{1\cdot 3\cdot 5\cdots(2n-1)\sqrt{n}}$$

Así que:

$$\left(\frac{2^{2n}(n!)^2}{(2n)!\sqrt{n}}\right)^2 = \frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdots(2n)2n)}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdots(2n-1)(2n-1)\cdot n} = \frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdots(2n)2n)}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdots(2n-1)(2n-1)(2\cdot n+1)}\frac{2n+1}{n}$$

Por lo tanto, utilizando la fórmula de Wallis, se obtiene:

$$\lim_{n\rightsquigarrow\infty}\left(\frac{2^{2n}(n!)^2}{(2n)!\sqrt{n}}\right)^2 = \lim_{n\rightsquigarrow\infty}\frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdots(2n)2n)}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdots(2n-1)(2n-1)(2\cdot n+1)}\frac{2n+1}{n} = \pi$$

de lo cual se sigue el resultado.

■

Lema 6.49. *Sea $n \in \mathbb{N}$, entonces* $\frac{1}{2}\ln(1+\frac{1}{n}) = \frac{1}{2n+1} + \frac{1}{3}\frac{1}{(2n+1)^3} + \frac{1}{5}\frac{1}{(2n+1)^5} + \cdots$

Demostración

Desarrollando la función $f(x) = \ln(1+x)$ en serie de Taylor alrededor del cero, se obtiene:

$$\ln(1+x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 + \cdots$$

$$\ln(1-x) = -x - \tfrac{1}{2}x^2 - \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 - \cdots$$

Así que:

$$\tfrac{1}{2}\ln\tfrac{1+x}{1-x} = \tfrac{1}{2}\ln(1+x) - \tfrac{1}{2}\ln(1-x) = x + \tfrac{1}{3}x^3 + \tfrac{1}{5}x^5 + \cdots$$

Tomando $x = \frac{1}{2n+1}$, se tiene $\frac{1+x}{1-x} = 1 + \frac{1}{n}$, así que:

$$\tfrac{1}{2}\ln(1+\tfrac{1}{n}) = \tfrac{1}{2n+1} + \tfrac{1}{3}\tfrac{1}{(2n+1)^3} + \tfrac{1}{5}\tfrac{1}{(2n+1)^5} + \cdots$$

Teorema 6.50 (**Fórmula de Stirling**). *Sea $n \in \mathbb{N}$, entonces $n! = \sqrt{2\pi}n^n e^{-n}\sqrt{n}e^{\theta_n}$, en donde $\frac{1}{12n+1} < \theta_n < \frac{1}{12n}$.*

Demostración

Sea $\varphi(n) = \frac{n!}{n^n e^{-n}\sqrt{n}}$, entonces:

$\frac{\varphi(n)}{\varphi(n+1)} = \frac{(n+1)^{n+1}e^{-1}\sqrt{n+1}}{(n+1)n^n\sqrt{n}} = \left(1+\frac{1}{n}\right)^{n+\frac{1}{2}} e^{-1}$

Así que, utilizando el lema 6.49:

$\ln\frac{\varphi(n)}{\varphi(n+1)} = \left(n+\frac{1}{2}\right)\ln\left(1+\frac{1}{n}\right) - 1$

$= (2n+1)\left(\frac{1}{2n+1} + \frac{1}{3}\frac{1}{(2n+1)^3} + \frac{1}{5}\frac{1}{(2n+1)^5} + \cdots\right) - 1$

$= \frac{1}{3}\frac{1}{(2n+1)^2} + \frac{1}{5}\frac{1}{(2n+1)^4} + \frac{1}{7}\frac{1}{(2n+1)^6} + \cdots > 0$

Por lo tanto, $\frac{\varphi(n)}{\varphi(n+1)} > 1$ para cualquier $n \in \mathbb{N}$, lo cual significa que la sucesión $\{\varphi(n)\}$ es monótona decreciente.

Además:

$\frac{1}{3}\frac{1}{(2n+1)^2} + \frac{1}{5}\frac{1}{(2n+1)^4} + \frac{1}{7}\frac{1}{(2n+1)^6} + \cdots < \frac{1}{3}\left[\frac{1}{(2n+1)^2} + \frac{1}{(2n+1)^4} + \frac{1}{(2n+1)^6} + \cdots\right]$

$= \frac{1}{3}\frac{1}{(2n+1)^2-1} = \frac{1}{12n(n+1)} = \frac{1}{12n} - \frac{1}{12(n+1)}$

Por lo tanto, $\ln\varphi(n) - \frac{1}{12n} < \ln\varphi(n+1) - \frac{1}{12(n+1)}$ para cualquier $n \in \mathbb{N}$, lo cual significa que la sucesión $\{\ln\varphi(n) - \frac{1}{12n}\}$ es monótona creciente y entonces también la sucesión $\{\varphi(n)e^{-\frac{1}{12n}}\}$ es monótona creciente.

También se tiene:

$\frac{1}{3}\frac{1}{(2n+1)^2} + \frac{1}{5}\frac{1}{(2n+1)^4} + \frac{1}{7}\frac{1}{(2n+1)^6} + \cdots > \frac{1}{3}\frac{1}{(2n+1)^2} > \frac{1}{12n+1} - \frac{1}{12(n+1)+1}$

Por lo tanto, $\ln\varphi(n) - \frac{1}{12n+1} > \ln\varphi(n+1) - \frac{1}{12(n+1)+1}$ para cualquier $n \in \mathbb{N}$, lo cual significa que la sucesión $\{\ln\varphi(n) - \frac{1}{12n+1}\}$ es monótona decreciente y entonces también la sucesión $\{\varphi(n)e^{-\frac{1}{12n+1}}\}$ es monótona decreciente.

Sea $A = \lim_{n\rightsquigarrow\infty}\varphi(n)$, entonces:

$\lim_{n\rightsquigarrow\infty}\varphi(n)e^{-\frac{1}{12n}} = A$ y, como la sucesión $\{\varphi(n)e^{-\frac{1}{12n}}\}$ es monótona creciente, $\varphi(n)e^{-\frac{1}{12n}} \leq A$ para cualquier $n \in \mathbb{N}$. En particular, $A > 0$.

$\lim_{n\rightsquigarrow\infty}\varphi(n)e^{-\frac{1}{12n+1}} = A$ y, como la sucesión $\{\varphi(n)e^{-\frac{1}{12n+1}}\}$ es monótona decreciente, $A \leq \varphi(n)e^{-\frac{1}{12n+1}}$ para cualquier $n \in \mathbb{N}$.

De las siguientes desigualdades:

$\varphi(n)e^{-\frac{1}{12n}} \leq A \leq \varphi(n)e^{-\frac{1}{12n+1}}$ para cualquier $n \in \mathbb{N}$

se sigue:

$Ae^{\frac{1}{12n+1}} \leq \varphi(n) \leq Ae^{\frac{1}{12n}}$ para cualquier $n \in \mathbb{N}$

Así que:

$\frac{n!}{n^n e^{-n}\sqrt{n}} = \varphi(n) = Ae^{\theta_n}$, en donde $\frac{1}{12n+1} \leq \theta_n \leq \frac{1}{12n}$ y $A = \text{lím}_{n\rightsquigarrow\infty} \varphi(n) > 0$.

Resta únicamente probar que $A = \sqrt{2\pi}$. Para esto, se tiene:

$\frac{\varphi^2(n)}{\varphi(2n)} = \frac{(2n)^{2n}e^{-2n}\sqrt{2n}}{(2n)!}\frac{(n!)^2}{n^{2n}e^{-2n}n} = \frac{2^{2n}(n!)^2\sqrt{2}}{(2n)!\sqrt{n}}$

Así que, por el corolario 6.48:

$A = \frac{A^2}{A} = \text{lím}_{n\rightsquigarrow\infty} \frac{\varphi^2(n)}{\varphi(2n)} = \text{lím}_{n\rightsquigarrow\infty} \frac{2^{2n}(n!)^2\sqrt{2}}{(2n)!\sqrt{n}} = \sqrt{2\pi}$

■

EJERCICIOS

Ejercicio 6.1. *Sea X una variable aleatoria continua. Exprese $P[|X-1|>2]$ en términos de la función de distribución, F_X, de X.*

Ejercicio 6.2. *Sea X una variable aleatoria con función de distribución dada por:*

$$F_X(x) = \begin{cases} 0 & si\ x < 0 \\ \frac{1}{2}x^2 & si\ x \in [0,1) \\ \frac{3}{4} & si\ x \in [1,2) \\ \frac{1}{4}(x+1) & si\ x \in [2,3) \\ 1 & si\ x \geq 3 \end{cases}$$

Encuentre a) $P\left[\frac{1}{2} < X < \frac{5}{2}\right]$, b) $P[X<1]$, c) $P[X=1]$ y d) $P\left[X=\frac{1}{2}\right]$.

Ejercicio 6.3. *Dé un ejemplo que una función de distribución discreta que sea estrictamente creciente.*

Ejercicio 6.4. *Sea f la función definida por:*

$$f(x) = \begin{cases} x+1 & si\ -1 < x < 0 \\ 2x-6 & si\ 3 < x < c \\ 0 & en\ otro\ caso \end{cases}$$

en donde c es una constante. a) Determine el valor de c de tal manera que f sea una función de densidad. b) Encuentre la función de distribución correspondiente a f.

Ejercicio 6.5. *Sea f la función definida por:*

$$f(x) = \begin{cases} C(x^3 - 2x) & \textit{si } 0 < x < \frac{5}{2} \\ 0 & \textit{en otro caso} \end{cases}$$

en donde C es una constante. ¿Podría ser f una función de densidad?

Ejercicio 6.6. *El volumen de ventas (en miles de galones) por semana de una estación de gasolina es una variable aleatoria con función de densidad dada por:*

$$f(x) = \begin{cases} 5(1 - x)^4 & \textit{si } 0 < x < 1 \\ 0 & \textit{en otro caso} \end{cases}$$

Si la estación es abastecida de gasolina una vez por semana, ¿cuál debe ser la capacidad del tanque para que la probabilidad de que se rebase sea menor que 0.01?

Ejercicio 6.7. *Una urna contiene n tarjetas numeradas del 1 al n. Un experimento aleatorio consiste en seleccionar tarjetas al azar, una a una y con reemplazo, hasta que se obtiene una tarjeta que ya se seleccionó con anterioridad. Encuentre la función de densidad de la variable aleatoria X definida como el número de tarjetas que son seleccionadas en este proceso.*

Ejercicio 6.8. *Muestre con un ejemplo que existen variables aleatorias X y Y tales que X^2 y Y^2 son independientes, pero X y Y no lo son.*

CAPÍTULO 7

VARIABLES ALEATORIAS DISCRETAS

Todo debería hacerse tan simple como sea posible, pero no más simple.

Albert Einstein

7.1. Distribución binomial

DEFINICIÓN 7.1 (**Distribución binomial**). *Si $n \in \mathbb{N}$ y $p \in [0,1]$, se dice que una variable aleatoria X tiene distribución binomial con parámetros n y p si su función de densidad está dada por la fórmula:*

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x \in \{0, \ldots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

El nombre binomial obedece a que los términos $\binom{n}{x} p^x (1-p)^{n-x}$ son los del desarrollo de $(p+q)^n$, en donde $q = 1 - p$.

La distribución binomial surgió prácticamente desde los inicios del Cálculo de Probabilidades, al considerarse el experimento aleatorio consistente en lanzar n veces un dado y buscar la probabilidad de obtener una de las caras cierto número de veces. Se encuentra por primera vez en el tratado sobre probabilidad de Jacques Bernoulli titulado "Ars Conjectandi", el cual fue publicado postumamente en el año 1713. Surge ahí como solución de un problema que plantea Christiaan Huygens en su libro, "De ratiociniis in ludo aleæ", el cual consiste en lo siguiente:

¿Cuántas veces debe lanzarse un dado para que sea más favorable obtener por lo menos dos veces el número 6?

Bernoulli planteó y resolvió la siguiente forma general del problema de Huygens:

Supongamos que se tiene un dado con $a = b + c$ caras y que un jugador apuesta a que, en n tiradas sucesivas del dado, obtendrá por lo menos m veces alguna de las b caras del dado, ¿cuál es la probabilidad de que tal jugador consiga su objetivo?

Bernoulli resolvió este problema estableciendo primero que, el lanzar n veces uno de tales dados es equivalente a lanzar una vez n dados con la misma configuración. Al lanzar n dados con a caras cada uno, hay a^n posibles resultados; de ellos hay c^n casos en los cuales no se obtiene alguna de las b caras con ninguno de los n dados; además, hay $b^k c^{n-k}$ casos en que, con cada uno de k dados particulares, se obtiene alguna de las b caras y en el resto se obtiene alguna de las c caras. Habiendo un total de $\binom{n}{k}$ maneras en que se pueden seleccionar k dados particulares de entre los n, obtuvo entonces que de los a^n posibles resultados del lanzamiento de los n dados, hay $\binom{n}{k} b^k c^{n-k}$ casos en los cuales el jugador obtiene k de las caras por las que ha apostado; la probabilidad de que éste no consiga su objetivo está dada entonces por:

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

De manera general, la distribución binomial se obtiene dentro del contexto de los llamados ensayos de Bernoulli (ver sección 5.1):

Supongamos que se realizan n ensayos de Bernoulli de tal manera que se satisfagan las siguientes condiciones:

- Cada ensayo es independiente de los demás.
- La probabilidad de éxito en cada ensayo es igual a una constante p.

Si definimos X como el número de éxitos que se obtienen en los n ensayos, entonces X tiene distribución binomial con parámetros n y p. En efecto, para $k \in \{1, \ldots, n\}$, la probabilidad de obtener k éxitos y $n-k$ fracasos, en lugares fijos, es igual a $p^k(1-p)^k$, y hay $\binom{n}{k}$ maneras en que se pueden seleccionar k lugares fijos para los éxitos, por lo tanto, la probabilidad de obtener exactamente k éxitos en los n ensayos está dada por $\binom{n}{k}p^k(1-p)^k$.

Como vimos en el ejemplo 3.9, esta situación se presenta de manera típica en problemas de muestreo con reemplazo. De manera específica, cuando se toma una muestra aleatoria con reemplazo de una población formada por dos tipos de objetos, I y II, la variable aleatoria X definida como el número de objetos de tipo I (o II), que se obtienen en la muestra, tiene distribución binomial.

Proposición 7.2. *Los términos $t_k = \binom{n}{k}p^k(1-p)^k$, de una distribución binomial, crecen con k hasta alcanzar su máximo valor cuando $np+p-1 \leq k \leq np+p$, después de lo cual decrecen con k.*

Demostración

Para $k \in \{0, \ldots, n\}$, sea $t_k = \binom{n}{k} p^k (1-p)^{n-k}$. Se tiene entonces, para $k \in \{1, \ldots, n\}$:

$$\frac{t_{k-1}}{t_k} = \frac{\binom{n}{k-1} p^{k-1} q^{n-k+1}}{\binom{n}{k} p^k q^{n-k}} = \frac{kq}{(n-k+1)p}$$

De manera que $t_{k-1} \leq t_k$ si y sólo si $kq \leq np - kp + p$, es decir, $k \leq np + p$.

Para $k \in \{0, \ldots, n-1\}$, se tiene:

$$\frac{t_k}{t_{k+1}} = \frac{\binom{n}{k} p^k q^{n-k}}{\binom{n}{k+1} p^{k+1} q^{n-k+1}} = \frac{(k+1)q}{(n-k)p}$$

De manera que $t_k \geq t_{k+1}$ si y sólo si $kq + q \geq np - kp$, es decir, $k \geq np - q = np + p - 1$.

■

A continuación se presentan las gráficas de algunas funciones de densidad tipo binomial. Con el objeto de visualizar más fácilmente la forma que adquieren, se han trazado continuas, graficando la función $f(x) = \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)} p^x (1-p)^{n-x}$, definida en el intervalo $[0, n]$. Evidentemente, los valores de la función de densidad corresponden únicamente a los valores enteros comprendidos entre 0 y n inclusive.

$n = 10$, $p = \frac{1}{8}$ $n = 10$, $p = \frac{1}{2}$ $n = 10$, $p = \frac{7}{8}$

$n = 100$, $p = \frac{1}{20}$ $n = 100$, $p = \frac{2}{5}$ $n = 100$, $p = \frac{19}{20}$

EJEMPLO 7.3. *Una parte de un satélite está compuesta por* 10 *componentes independientes, de tal manera que dicha parte funciona únicamente si por lo menos* 6 *de sus partes funcionan. En un día lluvioso cada una de las componentes funciona con probabilidad* $p_1 = \frac{1}{2}$*, mientras que en un día sin lluvia cada una de las componentes funciona con probabilidad* $p_2 = \frac{3}{4}$*. Suponiendo que la probabilidad de que mañana llueva es igual a* $\frac{1}{4}$*, ¿cuál es la probabilidad de que la mencionada parte del satélite funcione bien?*

Solución

Sea X *el número de componentes que funcionan y definamos los eventos:*

A*: la parte funciona.*

B*: el día es lluvioso.*

Se tiene entonces:

$P(A \mid B) = P\left(X \geq 6 \mid B\right) = \sum_{j=6}^{10} \binom{10}{j} p_1^j (1-p_1)^{10-j}$

$P(A \mid B^c) = P\left(X \geq 6 \mid B^c\right) = \sum_{j=6}^{10} \binom{10}{j} p_2^j (1-p_2)^{10-j}$

$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$

$= \frac{1}{4} \sum_{j=6}^{10} \binom{10}{j} p_1^j (1-p_1)^{10-j} + \frac{3}{4} \sum_{j=6}^{10} \binom{10}{j} p_2^j (1-p_2)^{10-j} = \frac{1647613}{2097152} = 0.78564$

Obsérvese que, para $i \neq j$*, los eventos, 'la componente* i *funciona' y 'la componente* j *funciona', no son independientes. En efecto, definamos los eventos:*

C_j*: la componente* j *funciona.*

Entonces:

$P(C_j) = P(C_j \mid B)P(B) + P(C_j \mid B^c)P(B^c) = \frac{1}{2}\frac{1}{4} + \frac{3}{4}\frac{3}{4} = \frac{11}{16}$

$P(C_i \cap C_j) = P(C_i \cap C_j \mid B)P(B) + P(C_i \cap C_j \mid B^c)P(B^c) = \frac{1}{4}\frac{1}{4} + \frac{9}{16}\frac{3}{4} = \frac{31}{64} = \frac{124}{256}$

$P(C_i)P(C_j) = \frac{11}{16}\frac{11}{16} = \frac{121}{256}$

Así que, $P(C_i \cap C_j) \neq P(C_i)P(C_j)$*.*

Esto se puede explicar observando que si C_1 *ocurre, entonces la probabilidad de lluvia disminuye, en efecto:*

$P(B) = \frac{1}{4} = \frac{11}{44}$

mientras que:

$P(B \mid C_1) = \frac{P(C_1|B)P(B)}{P(C_1)} = \frac{\frac{1}{2}\frac{1}{4}}{\frac{11}{16}} = \frac{2}{11} = \frac{8}{44}$

Esto hace que la probabilidad de que C_2 ocurra se incremente, en efecto:

$P(C_2) = \frac{11}{16} = \frac{121}{176}$

mientras que:

$P(C_2 \mid C_1) = \frac{P(C_{i1} \cap C_2)}{P(C_1)} = \frac{\frac{31}{64}}{\frac{11}{16}} = \frac{31}{44} = \frac{124}{176}$

▲

La distribución binomial es una de las distribuciones más importantes de la Teoría de la Probabilidad pues su estudio ha conducido a resultados fundamentales. Jacques Bernoulli, además de ser el primero en obtenerla, realizó un estudio a fondo de ella al plantearse el problema de demostrar que la frecuencia relativa de ocurrencia de un evento se aproxima a la probabilidad del mismo.

Si p es la probabilidad de ocurrencia de un evento A en un determinado experimento aleatorio, entonces llamando X_n al número de veces que ocurre A al realizar n repeticiones independientes del experimento, X_n tiene distribución binomial con parámetros n y p.

Ahora bien, $\frac{X_n}{n}$ es la frecuencia relativa de ocurrencia de A en las n repeticiones del experimento. Bernoulli demostró entonces que dado cualquier número $\varepsilon > 0$, se tiene:

$$\text{lím}_{n \rightsquigarrow \infty} P\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right] = 0$$

Este resultado, cuya demostración llevó a Bernoulli 20 años, se conoce como el teorema de Bernoulli y constituye el primero de los llamados **teoremas límite** de la Teoría de la Probabilidad. Su publicación marcó una línea de investigación, la cual se extendió durante un periodo de más de 200 años, hasta concluir en las versiones generales de este teorema y otros similares.

Dentro de la teoría moderna, el teorema de Bernoulli es únicamente un caso particular de la ley débil de los grandes números. Sin embargo, resulta ilustrativa una demostración directa, la cual puede realizarse con la herramienta que se tiene desarrollada hasta este momento.

PROPOSICIÓN 7.4 (**Teorema de Bernoulli**). *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 y ε un número positivo arbitrario, entonces:*

$$\text{lím}_{n \rightsquigarrow \infty} P\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right] = 0$$

Demostración

Sabemos que X_n tiene distribución binomial con parámetros n y p. Para $j \in \{0, \ldots, n\}$, sea $t_j = \binom{n}{j} p^j (1-p)^{n-j}$. De acuerdo con la proposición 7.2, los términos t_j crecen con j hasta alcanzar su máximo valor cuando $np + p - 1 \leq j \leq np + p$, después de lo cual decrecen con j. Además, la sucesión $s_j = \frac{t_j}{t_{j-1}}$, con $j \in \{1, \ldots, n\}$, es decreciente. En efecto, para $j \in \{2, \ldots, n\}$, se tiene:

$$s_j = \frac{t_j}{t_{j-1}} = \frac{(n-j+1)p}{jq} < \frac{(n-j+2)p}{(j-1)q} = \frac{t_{j-1}}{t_{j-2}} = s_{j-1}$$

Sea $j \geq k > (n+1)p$ y $r = \frac{t_k}{t_{k-1}} = \frac{(n-k+1)p}{kq}$, entonces $0 < r < 1$ y:

$$t_j = \frac{t_j}{t_{j-1}} t_{j-1} \leq \frac{t_k}{t_{k-1}} t_{j-1} = \frac{(n-k+1)p}{kq} t_{j-1} = r t_{j-1}$$

Así que:

$$t_j \leq r t_{j-1} \leq r^2 t_{j-2} \leq \cdots \leq r^{j-k} t_k$$

Por lo tanto:

$$P[X_n \geq k] = \sum_{j=k}^n t_j \leq t_k \sum_{j=k}^n r^{j-k} < \frac{1}{1-r} t_k = \frac{kq}{k-(n+1)p} t_k$$

Sea m el único número entero tal que $(n+1)p - 1 < m \leq (n+1)p$, entonces, como $t_m > t_{m+1} > \cdots > t_{k-1} > t_k$, se tiene:

$$1 > t_m + t_{m+1} + \cdots + t_{k-1} > (k-m)t_k \geq [k - (n+1)p]\, t_k$$

Así que $t_k < \frac{1}{k-(n+1)p}$. Por lo tanto:

$$P[X_n \geq k] < \frac{kq}{[k-(n+1)p]^2}$$

Dada $\varepsilon > 0$ y $n \in \mathbb{N}$ tal que $n\varepsilon > 1$, sea k_0 el único entero tal que $np + n\varepsilon < k_0 \leq np + n\varepsilon + 1$, entonces:

$$P\left[\frac{X_n}{n} - p > \varepsilon\right] = P[X_n > np + n\varepsilon] = P[X_n \geq k_0] < \frac{k_0 q}{[k_0-(n+1)p]^2}$$

$$\leq \frac{(np+n\varepsilon+1)q}{[np+n\varepsilon-(n+1)p]^2} = \frac{(np+n\varepsilon+1)q}{[n\varepsilon-p]^2}$$

Por lo tanto:

$\lim_{n\rightsquigarrow\infty} P\left[\frac{X_n}{n} - p > \varepsilon\right] = 0$

Sea Y_n el número de veces que no ocurre el evento A en las primeras n repeticiones del experimento, se tiene entonces $Y_n = n - X_n$, así que:

$\lim_{n\rightsquigarrow\infty} P\left[\frac{X_n}{n} - p < -\varepsilon\right] = \lim_{n\rightsquigarrow\infty} P\left[q - \frac{Y_n}{n} < -\varepsilon\right] = \lim_{n\rightsquigarrow\infty} P\left[\frac{Y_n}{n} - q > \varepsilon\right] = 0$

Por lo tanto:

$\lim_{n\rightsquigarrow\infty} P\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right] = 0$

■

El teorema de Bernoulli no asegura que, en una sucesión infinita de repeticiones del experimento $\mathcal{E}$, la frecuencia relativa de ocurrencia de A, en las primeras n repeticiones del experimento, tenga un límite igual a p, lo cual demostraría la validez de la interpretación frecuencial de la probabilidad. Lo que afirma es que, dada cualquier $\varepsilon > 0$, la **probabilidad** del evento $\left[\left|\frac{X_n}{n} - p\right| \leq \varepsilon\right]$ se acerca a 1 a medida que n crece. Es decir, establece una relación en términos de probabilidades. Para aclarar el resultado, supongamos que el experimento $\mathcal{E}$ consiste en el lanzamiento de un dado y que A representa la obtención de un 6 al realizar el experimento. La probabilidad $p = \frac{1}{6}$ significa entonces, en este caso, que se tiene 1 caso favorable y 5 desfavorables para la ocurrencia de A. En n repeticiones del experimento hay un total de 6^n posibles resultados, todos ellos con la misma probabilidad, de manera que si llamamos A_n (resp. B_n) al número de posibles resultados que favorecen la ocurrencia del evento $\left[\left|\frac{X_n}{n} - p\right| \leq \varepsilon\right]$ (resp. $\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right]$) en las n repeticiones del experimento, se tiene $A_n + B_n = 6^n$, $P\left[\left|\frac{X_n}{n} - p\right| \leq \varepsilon\right] = \frac{A_n}{6^n}$ y $P\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right] = \frac{B_n}{6^n}$. Así que:

$$\lim_{n\rightsquigarrow\infty} \frac{A_n}{B_n} = \lim_{n\rightsquigarrow\infty} \frac{P\left[\left|\frac{X_n}{n} - p\right| \leq \varepsilon\right]}{P\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right]} = \infty$$

Así que, en este caso, una formulación equivalente del teorema de Bernoulli consiste en decir que, a medida que n crece, el número de posibles resultados que favorecen la ocurrencia del evento $\left[\left|\frac{X_n}{n} - p\right| \leq \varepsilon\right]$ se hace mucho más grande que el número de posibles resultados que favorecen la ocurrencia del evento $\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right]$, de tal manera que su cociente tiende a ∞, pero no que el evento $\left[\left|\frac{X_n}{n} - p\right| \leq \varepsilon\right]$ ocurra con seguridad. De manera también equivalente, se puede decir que para determinar si el evento $B = \left[\left|\frac{X_n}{n} - p\right| \leq \varepsilon\right]$ ocurre, se pueden colocar A_n bolas blancas y B_n bolas negras en una urna, después de lo cual se selecciona una bola al azar. Si ésta resulta blanca, entonces B ocurre. Evidentemente, el que A_n sea muchísimo más grande que B_n (para n grande) no significa que al seleccionar una bola al azar de la urna, ésta sea necesariamente blanca. Tampoco significa que si se repite muchas veces el experimento aleatorio consistente en seleccionar, al azar y con reemplazo, una bola de esa urna, casi siempre se obtendrá una bola blanca. El que esto ocurriera sería un resultado experimental que reforzaría una interpretación frecuencial de la probabilidad, pero no una prueba de la validez del teorema de Bernoulli. De hecho, el teorema de Bernoulli, siendo un resultado teórico dentro del modelo de probabilidad que hemos definido, no requiere de ninguna verificación experimental, es válido como cualquier resultado que se demuestre en cualquier teoría matemática.

Por otro lado, si bien el teorema de Bernoulli no demuestra la validez de la interpretación frecuencial de la probabilidad, si se muestra con él que tal interpretación es compatible con el modelo probabilístico que hemos definido, haciendo así que las aplicaciones de la Teoría de la Probabilidad basadas en su interpretación frecuencial tengan bases más sólidas.

Además, la publicación del teorema de Bernoulli, en el año 1713, hizo renacer el interés por la Teoría de Probabilidad, la cual, después de la publicación del trabajo de Christiaan Huygens, en el año 1657, había quedado relegada, siendo vista únicamente como una curiosidad que tenía que ver exclusivamente con los juegos de azar. Siguiendo a Bernoulli vendrían después los trabajos de Abraham de Moivre, Siméon Denis Poisson y Pierre Simon Laplace, marcándose así, como dijimos arriba, una línea de investigación que se extendió por un periodo de más de 200 años.

7.2. Distribución geométrica

Definición 7.5 (**Distribución geométrica**). *Dado $p \in (0,1]$, se dice que una variable aleatoria X tiene distribución geométrica con parámetro p si su función de densidad está dada por la fórmula:*

$$f_X(x) = \begin{cases} p(1-p)^x & \text{si } x \in \{0,1,\ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

El nombre geométrica obedece a que los términos $p(1-p)^x$ son los términos de la serie geométrica $\sum_x p(1-p)^x$.

Como $\sum_{x=0}^{\infty} p(1-p)^x = 1$, efectivamente X resulta ser una variable aleatoria discreta.

La distribución geométrica surge también dentro del contexto de los ensayos de Bernoulli. Supongamos que se realiza una sucesión indefinida de ensayos de Bernoulli independientes, en cada uno de los cuales la probabilidad de éxito es igual a p. Definamos X como el número de fracasos que se obtienen antes del primer éxito. X tiene entonces distribución geométrica con parámetro p. En efecto, para $k \in \{0,1,\ldots\}$, definamos los siguientes eventos:

A_k: Se obtiene fracaso en cada uno de los primeros k ensayos.

B: Se obtiene éxito en el ensayo número k.

Entonces:

$P[X = k] = P(A_k \cap B) = P(A_k)P(B) = (1 - p)^k p$

A continuación se presentan las gráficas de algunas funciones de densidad tipo geométrica. Con el objeto de visualizar más fácilmente la forma que adquieren, se han trazado continuas, graficando la función $f(x) = p(1 - p)^x$, definida en el intervalo $[0, \infty)$. Evidentemente, los valores de la función de densidad corresponden únicamente a los valores enteros no negativos.

$p = \frac{1}{8}$ $\qquad$ $p = \frac{1}{2}$ $\qquad$ $p = \frac{7}{8}$

Ejemplo 7.6. *Cada una de* 12 *personas lanza una moneda indefinidamente, deteniendo sus lanzamientos en el momento en que obtiene por primera vez águila. ¿Cuál es la probabilidad de que por lo menos* 10 *de ellas realicen menos de* 4 *lanzamientos?*

Solución

Sea X el número de lanzamientos que se realizan antes de obtener por primera vez águila al lanzar una moneda indefinidamente y Y el número de personas que realizan menos de 4 *lanzamientos. Y tiene entonces distribución binomial con parámetros $n = 12$ y $p = P[X < 3]$. A su vez, X tiene distribución geométrica con parámetro $\frac{1}{2}$. Así que:*

$p = P[X < 3] = P[X \leq 2] = \sum_{k=0}^{2} \frac{1}{2} \left(\frac{1}{2}\right)^k = \frac{7}{8}$

$P[Y \geq 10] = \sum_{k=10}^{12} \binom{12}{k} \left(\frac{7}{8}\right)^k \left(\frac{1}{8}\right)^{12-k} = 0.818$

Ejemplo 7.7. *Sea X una variable aleatoria con distribución geométrica. Demuestre que para cualquier par de números enteros no negativos n y k, se tiene:*

$$P[X = k + n \mid X > n - 1] = P[X = k]$$

Obsérvese que este resultado dice que si en una sucesión de ensayos de Bernoulli independientes ya han ocurrido n fracasos, entonces la probabilidad de que haya k fracasos más antes del primer éxito se puede calcular como si se reiniciara el proceso. Este resultado se puede interpretar diciendo que una variable aleatoria con distribución geométrica no tiene memoria.

Solución

$P[X = k + n \mid X > n - 1] = \frac{P[X=k+n, X>n-1]}{P[X>n-1]} = \frac{P[X=k+n]}{P[X>n-1]}$

$= \frac{p(1-p)^{k+n}}{(1-p)^n} = p(1-p)^k = P[X = k]$

▲

Dado que una variable aleatoria X con distribución geométrica es discreta, se tiene $P[X < \infty] = 1$ (ver proposición 6.28). Esta propiedad aparentemente tan simple expresa que, en una sucesión indefinida de ensayos de Bernoulli independientes, todos con la misma probabilidad de éxito, la probabilidad de que en algún momento se produzca éxito es igual a 1. Como ya se expuso en la sección 5.3, la aplicación de este resultado produce otros que pueden parecer sorprendentes, por ejemplo, ahí vimos que si C es el conjunto de caracteres tipográficos (incluyendo el espacio blanco) que se utilizan en la obra Hamlet de Shakespeare y consideramos el experimento aleatorio consistente en escribir una secuencia indefinida de caracteres, cada uno seleccionado al azar del conjunto C, entonces la probabilidad de que en algún momento una subsecuencia de caracteres coincida exactamente con la obra de Shakespeare es igual a 1.

7.3. Distribución binomial negativa

DEFINICIÓN 7.8 (**Distribución binomial negativa**). *Si $r \in (0, \infty)$ y $p \in (0, 1]$, se dice que una variable aleatoria X tiene distribución binomial negativa con parámetros r y p si su función de densidad está dada por la fórmula:*

$$f_X(x) = \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x & \text{si } x \in \{0, 1, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

De acuerdo con las proposiciones 3.12 y 6.46, esta densidad puede escribirse también en la forma siguiente:

$$f_X(x) = \begin{cases} (-1)^x \binom{-r}{x} p^r (1-p)^x & \text{si } x \in \{0, 1, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{\Gamma(r+x)}{x!\Gamma(r)} p^r (1-p)^x & \text{si } x \in \{0, 1, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

El nombre binomial negativa obedece a que los términos $\binom{r+x-1}{x} p^r (1-p)^x$ son los del desarrollo de $(\frac{1}{p} - \frac{q}{p})^{-r}$, en donde $q = 1 - p$. En efecto, por el corolario 3.15 y la proposición 3.12, se tiene:

$$(\tfrac{1}{p} - \tfrac{q}{p})^{-r} = (-\tfrac{q}{p} + \tfrac{1}{p})^{-r} = \sum_{k=0}^{\infty} \binom{-r}{k} \left(-\tfrac{q}{p}\right)^k \left(\tfrac{1}{p}\right)^{-r-k}$$

$$= \sum_{k=0}^{\infty} (-1)^k \binom{-r}{k} p^r q^k = \sum_{k=0}^{\infty} \binom{r+k-1}{k} p^r q^k$$

La distribución binomial negativa se obtuvo por primera vez en las soluciones que dieron Blaise Pascal y Pierre de Fermat, en el año 1654, al llamado problema de la división de apuestas, el cual consiste en lo 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 y quien la gane acumula un punto, de tal manera que 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?

Para que se vea más claramente como surge la distribución binomial negativa en la solución de este problema, consideremos un juego que se interrumpe cuando a los jugadores A y B les faltan r y s puntos para ganar, respectivamente.

Partiendo de esa situación, el jugador A ganará el juego cuando, en las siguientes $r + s - 1$ partidas, B gane a lo más $s - 1$ de ellas, pero la probabilidad de que B gane k de $r+s-1$ partidas está dada por $\binom{r+s-1}{k} \frac{1}{2^{r+s-1}}$, es decir por una distribución binomial negativa. De esta forma, la probabilidad de que A gane el juego está dada por:

$$P(A) = \binom{r+s-1}{0} \frac{1}{2^{r+s-1}} + \binom{r+s-1}{1} \frac{1}{2^{r+s-1}} + \cdots + \binom{r+s-1}{s-1} \frac{1}{2^{r+s-1}}$$

Esencialmente, este resultado es el que encontraron tanto Pascal como Fermat.

La distribución binomial negativa se obtiene también dentro del contexto de los ensayos de Bernoulli. Supongamos que se realiza una sucesión indefinida de ensayos de Bernoulli independientes, en cada uno de los cuales la probabilidad de éxito es igual a p. Definamos X como el número de fracasos que se obtienen antes del éxito r. X tiene entonces distribución binomial negativa con parámetros r y p; en efecto, para $k \in \{0, 1, \ldots\}$, definamos los siguientes eventos:

A_k: Se obtienen $r-1$ éxitos en los primeros $r+k-1$ ensayos.

B: Se obtiene éxito en el ensayo número $r+k$.

Entonces:

$P[X=k] = P(A_k \cap B) = P(A_k)P(B) = \left[\binom{r+k-1}{r-1}p^{r-1}(1-p)^k\right] p = \binom{r+k-1}{k}p^r(1-p)^k$

Obsérvese que en el problema de la división de apuestas, partiendo de que a los jugadores A y B les faltan r y s puntos para ganar, respectivamente, el jugador A ganará el juego cuando, en las siguientes $r+s-1$ partidas, el número de derrotas antes del r-simo triunfo sea menor que s.

Proposición 7.9. *Los términos $t_k = \binom{r+k-1}{k}p^r(1-p)^k$, de una distribución binomial negativa, crecen con k hasta alcanzar su máximo valor cuando $\frac{(r-1)q}{p} - 1 \leq k \leq \frac{(r-1)q}{p}$, después de lo cual decrecen con k.*

Demostración

Para $k \in \{0, 1, \ldots\}$, sea $t_k = \binom{r+k-1}{k}p^r(1-p)^k$. Se tiene entonces, para $k \in \{1, 2, \ldots\}$:

$\frac{t_{k-1}}{t_k} = \frac{\binom{r+k-2}{k-1}p^r(1-p)^{k-1}}{\binom{r+k-1}{k}p^r(1-p)^k} = \frac{k}{(r+k-1)(1-p)}$

De manera que $t_{k-1} \leq t_k$ si y sólo si $k \leq (r+k-1)(1-p)$, es decir, $k \leq \frac{(r-1)q}{p}$.

Para $k \in \{0, 1, \ldots\}$, se tiene:

$\frac{t_k}{t_{k+1}} = \frac{\binom{r+k-1}{k}p^r(1-p)^k}{\binom{r+k}{k+1}p^r(1-p)^{k+1}} = \frac{k+1}{(r+k)(1-p)}$

De manera que $t_k \geq t_{k+1}$ si y sólo si $k+1 \geq (r+k)(1-p)$, es decir, $k \geq \frac{rq-1}{p} = \frac{(r-1)q}{p} - 1$.

■

A continuación se presentan las gráficas de algunas funciones de densidad tipo binomial negativa. Con el objeto de visualizar más fácilmente la forma que adquieren, se han trazado continuas, graficando la función $f(x) = \frac{\Gamma(r+x)}{\Gamma(x+1)\Gamma(r)}p^r(1-p)^x$, definida en el intervalo $[0, \infty)$. Evidentemente, los valores de la función de densidad corresponden únicamente a los valores enteros no negativos.

$r = 10$, $p = \frac{1}{8}$ $r = 10$, $p = \frac{1}{2}$ $r = 10$, $p = \frac{7}{8}$

$r = 100$, $p = \frac{1}{20}$ $r = 100$, $p = \frac{2}{5}$ $r = 100$, $p = \frac{19}{20}$

EJEMPLO 7.10. *Cuando una determinada máquina no está bien ajustada, la probabilidad de que produzca un artículo defectuoso es igual a* 0.2*. Cada día la máquina se pone a funcionar y se detiene para su ajuste en el momento en que ya ha producido* 3 *artículos defectuosos. ¿Cuál es la probabilidad de que, estando desajustada, la máquina producirá por lo menos* 5 *artículos antes de ser detenida para su ajuste?*

Solución

Sea X *el número de artículos que produce una máquina desajustada antes de ser detenida para su ajuste; los posible valores de* X *son entonces* $3, 4, \ldots$*. Si llamamos éxito a la producción de un artículo defectuoso, entonces, para* $k \in \{3, 4, \ldots\}$*, el evento* $[X = k]$ *ocurre cuando se tienen* $k - 3$ *fracasos antes del tercer éxito, por lo tanto:*

$$P[X = k] = \binom{k-1}{k-3}(0.8)^{k-3}(0.2)^3 = \binom{k-1}{2}(0.8)^{k-3}(0.2)^3$$

Así que:

$$P[X \geq 5] = 1 - P[X = 3] - P[X = 4]$$

$$= 1 - \binom{2}{2}(0.8)^0(0.2)^3 - \binom{3}{2}(0.8)^1(0.2)^3 = 0.9728$$

EJEMPLO 7.11. *Se lanza una moneda tantas veces como sea necesario hasta obtener 5 veces sol. Sabiendo que en los primeros 6 lanzamientos aún no se logra el objetivo,*

¿cuál es la probabilidad de que se requieran, en total, menos de 9 lanzamientos para lograrlo?

Solución

Llamemos éxito al hecho de obtener sol al lanzar la moneda una vez y sea X el número de fracasos antes del quinto éxito. X tiene entonces distribución binomial negativa con parámetros $p = \frac{1}{2}$ y $r = 5$, y se busca $P[X \leq 3 \mid X \geq 2]$.

$P[X \leq 3 \mid X \geq 2] = \frac{P[X \leq 3, X \geq 2]}{P[X \geq 2]} = \frac{P[X=2]+P[X=3]}{1-P[X=0]-P[X=1]} = \frac{\binom{6}{2}p^5(1-p)^2+\binom{7}{3}p^5(1-p)^3}{1-\binom{4}{0}p^5(1-p)^0-\binom{5}{1}p^5(1-p)}$

$= \frac{65}{228} = 0.285088$

▲

Dado que una variable aleatoria X con distribución binomial negativa es discreta, se tiene $P[X < \infty] = 1$ (ver proposición 6.28). Esta propiedad expresa que, para cualquier $r \in \mathbb{N}$, en una sucesión indefinida de ensayos de Bernoulli independientes, todos con la misma probabilidad de éxito, la probabilidad de obtener por lo menos r éxitos es igual a 1.

De aquí se sigue un resultado más fuerte, ya expuesto en la sección 5.3: Consideremos una sucesión indefinida de ensayos de Bernoulli independientes, todos con la misma probabilidad de éxito, y definamos los eventos:

B: se produce un número infinito de éxitos.

B_k: se producen por lo menos k éxitos.

Entonces, la sucesión de eventos $B_1, B_2, \ldots$ es monótona decreciente y su intersección es B. Se tiene entonces:

$P(B) = P\left(\bigcap_{k=1}^{\infty} B_k\right) = \text{lím}_{k \rightsquigarrow \infty} P(B_k) = 1$

Por lo tanto, con probabilidad 1, se produce una infinidad de éxitos.

Como ya se expuso en la sección 5.3, la aplicación de este resultado produce otros que pueden parecer sorprendentes, por ejemplo, ahí vimos que si C es el conjunto de caracteres tipográficos (incluyendo el espacio blanco) que se utilizan en la obra Hamlet de Shakespeare y consideramos el experimento aleatorio consistente en escribir una secuencia indefinida de caracteres, cada uno seleccionado al azar del conjunto C, entonces la probabilidad de que alguna subsecuencia de caracteres coincida exactamente con la obra de Shakespeare, *una infinidad de veces*, es igual a 1.

Para $r \in \mathbb{N}$, la distribución binomial negativa puede obtenerse de la distribución binomial observando que si X tiene distribución binomial negativa con parámetros r y p, entonces, para cada entero no negativo k, se tiene:

$P[X = k] = \binom{r+k-1}{r-1} p^r (1-p)^k = \frac{r}{r+k}\binom{r+k}{r} p^r (1-p)^k = \frac{r}{r+k} P[Y = r]$

en donde Y es una variable aleatoria con distribución binomial de parámetros $r + k$ y p.

7.4. Distribución Poisson

Definición 7.12 (**Distribución Poisson**). *Si λ es un número real positivo, se dice que una variable aleatoria X tiene distribución Poisson con parámetro λ si su función de densidad está dada por la fórmula:*

$$f_X(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{si } x \in \{0, 1, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

Ejemplo 7.13. *Supongamos que el número de accidentes que ocurren en una cierta carretera tiene distribución Poisson con parámetro $\lambda = 2$. ¿Cuál es la probabilidad de que en un día ocurran, en esa carretera, más de dos accidentes, sabiendo que ocurre por lo menos uno?*

Solución

Sea X el número de accidentes que ocurren en la carretera, entonces:

$P[X > 2 \mid X \geq 1] = \frac{P[X>2, X\geq 1]}{P[X\geq 1]} = \frac{P[X>2]}{P[X\geq 1]} = \frac{1-P[X\leq 2]}{1-P[X=0]} = \frac{1-e^{-\lambda}(1+\lambda+\lambda^2/2)}{1-e^{-\lambda}}$

$= \frac{1-5e^{-2}}{1-e^{-2}} = 0.37393$

Proposición 7.14. *Los términos $t_k = \frac{\lambda^k e^{-\lambda}}{k!}$, de una distribución Poisson, crecen con k hasta alcanzar su máximo valor cuando $\lambda - 1 \leq k \leq \lambda$, después de lo cual decrecen con k.*

Demostración

Para $k \in \{0, 1, \ldots\}$, sea $t_k = \frac{\lambda^k e^{-\lambda}}{k!}$. Se tiene entonces, para $k \in \{1, 2, \ldots\}$:

$\frac{t_{k-1}}{t_k} = \frac{k}{\lambda}$

De manera que $t_{k-1} \leq t_k$ si y sólo si $k \leq \lambda$.

Para $k \in \{0, 1, \ldots\}$, se tiene:

$\frac{t_k}{t_{k+1}} = \frac{k+1}{\lambda}$

De manera que $t_k \geq t_{k+1}$ si y sólo si $k + 1 \geq \lambda$, es decir, $k \geq \lambda - 1$.

■

A continuación se presentan las gráficas de algunas funciones de densidad tipo Poisson. Con el objeto de visualizar más fácilmente la forma que adquieren, se han trazado continuas, graficando la función $f(x) = \frac{\lambda^x e^{-\lambda}}{\Gamma(x+1)}$, definida en el intervalo $[0, \infty)$. Evidentemente, los valores de la función de densidad corresponden únicamente a los valores enteros no negativos.

$\lambda = \frac{5}{4}$ $\lambda = 5$ $\lambda = 95$

El nombre Poisson obedece a que fue Siméon Denis Poisson quien encontró esta distribución al demostrar que se obtiene como límite de una distribución binomial, resultado que fue publicado en 1837 en un artículo titulado "Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des regles générales du Calcul des Probabilités".

Teorema 7.15 (**Teorema de Poisson**). *Supongamos que, para cada $n \in \mathbb{N}$, X_n es una variable aleatoria con distribución binomial de parámetros n y $p \in (0,1)$ de tal manera que $\lambda = np$ es constante, entonces, para cualquier $k \in \{0, 1, \ldots\}$, se tiene:*

$$\text{lím}_{n \rightsquigarrow \infty} P[X_n = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

Demostración

$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}$

$= \frac{\lambda^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-k} \left(1 - \frac{\lambda}{n}\right)^n$

Así que:

$\text{lím}_{n \rightsquigarrow \infty} P[X = k] = \frac{\lambda^k}{k!} \text{lím}_{n \rightsquigarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = \frac{\lambda^k e^{-\lambda}}{k!}$

■

Con el objeto de comparar los valores exactos de una función de densidad binomial con sus valores aproximados por medio de una densidad Poisson, a continuación se muestran las gráficas que se obtienen con ambos valores. Con el objeto de apreciar mejor la aproximación, las gráfica de los valores exactos y aproximados de la densidad binomial $f_X(x)$ se han trazado continuas, expresándolos como $\frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}p^x(1-p)^{n-x}$ y $\frac{\lambda^x e^{-\lambda}}{\Gamma(x+1)}$, respectivamente. La línea sólida corresponde a los valores exactos, mientras que la línea punteada corresponde a los valores aproximados.

$n = 10$, $p = \frac{7}{8}$ $n = 10$, $p = \frac{1}{8}$ $n = 100$, $p = \frac{1}{20}$

Como se muestra a continuación, la distribución Poisson aproxima otras distribuciones.

TEOREMA 7.16. *Supongamos que, para cada $r \in \mathbb{N}$, X_n es una variable aleatoria con distribución binomial negativa de parámetros r y $p \in (0,1)$ de tal manera que $\lambda = r\frac{q}{p}$ es constante, entonces, para cualquier $k \in \{0, 1, \ldots\}$, se tiene:*

$$\text{lím}_{n \rightsquigarrow \infty} P[X_n = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

Demostración

$P[X = k] = \binom{r+k-1}{k} p^r (1-p)^k = \frac{(r+k-1)(r+k-2)\cdots(r)}{k!}(1 - \frac{\lambda}{r+\lambda})^r \frac{\lambda^k}{(r+\lambda)^k}$

$= \frac{\lambda^k}{k!} \frac{r+k-1}{r+\lambda} \frac{r+k-2}{r+\lambda} \cdots \frac{r}{r+\lambda}(1 - \frac{\lambda}{r+\lambda})^r$

Así que:

$\text{lím}_{r \rightsquigarrow \infty} P[X = k] = \frac{\lambda^k}{k!} \text{lím}_{r \rightsquigarrow \infty}(1 - \frac{\lambda}{r+\lambda})^r = \frac{\lambda^k}{k!} \text{lím}_{x \rightsquigarrow \infty}(1 - \frac{\lambda}{x})^{x-\lambda}$

$= \frac{\lambda^k}{k!} \text{lím}_{x \rightsquigarrow \infty}(1 - \frac{\lambda}{x})^x (1 - \frac{\lambda}{x})^{-\lambda} = \frac{\lambda^k e^{-\lambda}}{k!}$

■

Con el objeto de comparar los valores exactos de una función de densidad binomial negativa con sus valores aproximados por medio de una densidad Poisson, a continuación se muestran las gráficas que se obtienen con ambos valores. Con el objeto de apreciar mejor la aproximación, las gráfica de los valores exactos y aproximados de la densidad binomial negativa $f_X(x)$ se han trazado continuas, expresándolos como $\frac{\Gamma(r+x)}{\Gamma(x+1)\Gamma(r)}p^r(1-p)^x$ y $\frac{\lambda^x e^{-\lambda}}{\Gamma(x+1)}$, respectivamente. La línea sólida corresponde a los valores exactos, mientras que la línea punteada corresponde a los valores aproximados.

$r = 10$, $q = \frac{7}{8}$ $\quad$ $r = 10$, $q = \frac{1}{8}$ $\quad$ $r = 100$, $q = \frac{1}{20}$

Ejemplo 7.17. *Se prepara masa para hacer N galletas. ¿Cuántas pasas se deben de mezclar en la masa para evitar que, en promedio, en cada 100 galletas haya más de una sin pasas?*

Solución

Sea n el número de pasas que se deben de agregar a la masa. Como las pasas se mezclan en la masa, se puede asumir que cada pasa puede quedar en cualquiera de las N galletas. De manera que si consideramos una galleta y una pasa determinadas y llamamos éxito al hecho de que esa pasa quede en la galleta, la probabilidad de éxito es igual a $\frac{1}{N}$. Considerando las n pasas y una galleta determinada, llamando X al número de pasas que quedan en esa galleta, se tiene:

$$P[X = k] = \binom{n}{k}\left(\frac{1}{N}\right)^k\left(\frac{N-1}{N}\right)^{n-k} \approx \frac{\left(\frac{n}{N}\right)^k e^{-\frac{n}{N}}}{k!}$$

Se busca n tal que $P[X = 0] \leq 0.01$, es decir, $e^{-\frac{n}{N}} \leq 0.01$. Así que:

$$n \geq -N\ln(0.01) = 4.605N$$

En general, si se quiere evitar que, en promedio, en cada m galletas haya más de una sin pasas, se debe de tener $e^{-\frac{n}{N}} \leq \frac{1}{m}$, es decir, $n \geq -N\ln(m)$.

▲

La distribución Poisson se aplica de manera importante para el estudio de eventos que ocurren aleatoriamente en el tiempo.

De manera más específica, consideremos un experimento aleatorio consistente en la observación de eventos que ocurren aleatoriamente en el tiempo. Para cada intervalo I formado por números reales no negativos, sea N_I el número de eventos que ocurren en el intervalo de tiempo I y definamos $N_t = N_{[0,t]}$. Supongamos ahora que se satisfacen las siguientes condiciones:

(i) $N_0 = 0$

(ii) Para cualquier intervalo I, la distribución de N_I depende únicamente de la longitud de I.

(iii) Si $I_1, \ldots, I_n$ son intervalos ajenos y $k_1, \ldots, k_n$ son números enteros no negativos cualesquiera, entonces los eventos $[N_{I_1} = k_1], \ldots, [N_{I_n} = k_n]$ son independientes .

(iv) $P[N_t = 1]$ es, en una vecindad de $t = 0$, de un orden de magnitud comparable con t de tal manera que $\text{lím}_{t \rightsquigarrow 0} \frac{1}{t} P[N_t = 1]$ existe y
$\lambda = \text{lím}_{t \rightsquigarrow 0} \frac{1}{t} P[N_t = 1] > 0$

(v) $P[N_t \geq 2]$ es pequeña en una vecindad de $t = 0$ de tal manera que:
$\text{lím}_{t \rightsquigarrow 0} \frac{1}{t} P[N_t \geq 2] = 0$

Para k entero no negativo y $t > 0$, sea $P_k(t) = P[N_t = k]$. Además, para cada $n \in \mathbb{N}$, definamos $h = \frac{t}{n}$ y consideremos los intervalos $I_1 = [0, h)$, $I_2 = [h, 2h)$, $\ldots$, $I_{n-1} = [(n-2)h, (n-1)h)$, $I_n = [(n-1)h, t]$. Se tiene entonces, para cualquier $n \in \mathbb{N}$:

$P_k(t) = P[N_t = k] = P\left[N_t = k, N_{I_j} \leq 1 \text{ para cualquier } j \in \{1, \ldots, n\}\ \right]$

$+P[N_t = k, N_{I_j} \geq 2 \text{ para alguna } j \in \{1, \ldots, n\}]$

Pero:

$P[N_t = k, N_{I_j} \geq 2 \text{ para alguna } j \in \{1, \ldots, n\}]$

$\leq P[N_t = k, N_{I_j} \geq 2 \text{ para alguna } j \in \{1, \ldots, n\}]$

$= P[\bigcup_{j=1}^n \left[N_{I_j} \geq 2\right]] \leq \sum_{j=1}^n P\left[N_{I_j} \geq 2\right] = nP[N_h \geq 2] = \frac{t}{h} P[N_h \geq 2]$

Así que:

$\text{lím}_{n \rightsquigarrow \infty} P[N_t = k, N_{I_j} \geq 2 \text{ para alguna } j \in \{1, \ldots, n\}] \leq \text{lím}_{n \rightsquigarrow \infty} \frac{t}{h} P[N_h \geq 2] = 0$

Además:

$P\left[N_t = k, N_{I_j} \leq 1 \text{ para cualquier } j \in \{1, \ldots, n\}\ \right]$

$= \binom{n}{k} \left(P[N_h = 1]\right)^k \left(P[N_h = 0]\right)^{n-k}$

$= \binom{n}{k} \left(P\left[N_h = 1\right]\right)^k \left(1 - P\left[N_h = 1\right] - P\left[N_h \geq 2\right]\right)^{n-k}$

$= \frac{n(n-1)\cdots(n-k+1)}{k!} \frac{(\lambda t)^k}{n^k} \frac{n^k}{(\lambda t)^k} \left(P\left[N_h = 1\right]\right)^k \left(1 - P\left[N_h = 1\right] - P\left[N_h \geq 2\right]\right)^{n-k}$

$= \frac{(\lambda t)^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \frac{1}{\lambda^{-k}} \left(\frac{1}{h} P\left[N_h = 1\right]\right)^k$

$\left(1 - P\left[N_h = 1\right] - P\left[N_h \geq 2\right]\right)^k \left(1 - P\left[N_h = 1\right] - P\left[N_h \geq 2\right]\right)^n$

Pero:

$\lim_{n \rightsquigarrow \infty} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) = 1$

$\lim_{n \rightsquigarrow \infty} \frac{1}{\lambda^k} \left(\frac{1}{h} P\left[N_h = 1\right]\right)^k = 1$

$\lim_{n \rightsquigarrow \infty} \left(1 - P\left[N_h = 1\right] - P\left[N_h \geq 2\right]\right)^k = 1$

$\lim_{n \rightsquigarrow \infty} \left(1 - P\left[N_h = 1\right] - P\left[N_h \geq 2\right]\right)^n$

$= \lim_{n \rightsquigarrow \infty} \left(1 - \frac{1}{n} \left\{\frac{t}{h} P\left[N_h = 1\right] - \frac{t}{h} P\left[N_h \geq 2\right]\right\}\right)^n = \lim_{n \rightsquigarrow \infty} \left(1 - \frac{1}{n} f(n)\right)^n$

en donde $\lim_{n \rightsquigarrow \infty} f(n) = \lambda t$.

Por lo tanto, utilizando el lema 7.18, se obtiene:

$\lim_{n \rightsquigarrow \infty} P\left[N_t = k, N_{I_j} \leq 1 \text{ para cualquier } j \in \{1, \ldots, n\}\ \right] = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$

Y entonces:

$P_k(t) = P[N_t = k] = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$

LEMA 7.18. *Sea* (s_n) *una sucesión convergente y* $s = \lim_{n \rightsquigarrow \infty} s_n$*, entonces:*

$$\lim_{n \rightsquigarrow \infty} \left(1 - \frac{1}{n} s_n\right)^n = e^{-s}$$

Demostración

Dada $\varepsilon > 0$, sea N tal que, para cualquier $n \geq N$ se tenga $s - \varepsilon < s_n < s + \varepsilon$, así como $1 - \frac{1}{n}(s + \varepsilon) > 0$, $1 - \frac{1}{n}(s - \varepsilon) > 0$ y $1 - \frac{1}{n} s_n > 0$. Entonces, para $n \geq N$, se tiene:

$\left[1 - \frac{1}{n}(s + \varepsilon)\right]^n < \left[1 - \frac{1}{n} s_n\right]^n < \left[1 - \frac{1}{n}(s - \varepsilon)\right]^n$

Por lo tanto:

$\liminf_{n \rightsquigarrow \infty} \left[1 - \frac{1}{n} s_n\right]^n \geq \lim_{n \rightsquigarrow \infty} \left[1 - \frac{1}{n}(s + \varepsilon)\right]^n = e^{-(s+\varepsilon)}$

$\limsup_{n\rightsquigarrow\infty}\left[1-\frac{1}{n}s_n\right]^n \leq \lim_{n\rightsquigarrow\infty}\left[1-\frac{1}{n}(s-\varepsilon)\right]^n = e^{-(s-\varepsilon)}$

De manera que, como esto vale para cualquier $\varepsilon > 0$, se tiene:

$e^{-s} \leq \liminf_{n\rightsquigarrow\infty}\left[1-\frac{1}{n}s_n\right]^n \leq \limsup_{n\rightsquigarrow\infty}\left[1-\frac{1}{n}s_n\right]^n \leq e^{-s}$

Así que:

$\lim_{n\rightsquigarrow\infty}\left[1-\frac{1}{n}s_n\right]^n = e^{-s}$

■

La familia de variables aleatorias $(N_t)_{t\geq 0}$ forman lo que se conoce como un proceso de Poisson, de acuerdo con la siguiente definición:

Definición 7.19 (**Proceso de Poisson**). *Se dice que una familia $\{P_t : t \geq 0\}$ de variables aleatorias discretas forma un proceso de Poisson de parámetro λ si se satisfacen las siguientes propiedades:*

(i) $P_0 = 0$
(ii) *Si $0 < t_1 < t_2 < \cdots < t_n$, entonces las variables aleatorias P_{t_1}, $P_{t_2} - P_{t_1}$, $P_{t_3} - P_{t_2}$, ..., $P_{t_n} - P_{t_{n-1}}$ son independientes.*
(iii) *Si $s < t$, entonces la variable aleatoria $P_t - P_s$ tiene distribución Poisson con parámetro $\lambda(t-s)$.*

Los siguientes son ejemplos de eventos que ocurren aleatoriamente en el tiempo y que satisfacen aproximadamente las condiciones *i-v*:

(i) Emisión de partículas α por una fuente radioactiva.
(ii) Intercambio de cromosomas entre células inducido por irradiación de rayos X.
(iii) Surgimiento, por reproducción, de colonias de bacterias en una determinada región.
(iv) Llamadas que llegan a una central telefónica.
(v) Errores de imprenta en un libro.

7.5. Distribución hipergeométrica

Definición 7.20 (**Distribución hipergeométrica**). *Si r, s y n son números enteros positivos tales que $n \leq r+s$, se dice que una variable aleatoria X tiene distribución hipergeométrica con parámetros r, s y n si su función de densidad está dada por la fórmula:*

$$f_X(x) = \begin{cases} \frac{\binom{r}{x}\binom{s}{n-x}}{\binom{r+s}{n}} & si\ x \in \{m, \dots, M\} \\ 0 & en\ otro\ caso \end{cases}$$

en donde $m = \text{máx}\{0, n-s\}$ *y* $M = \text{mín}\{n, r\}$.

Obsérvese que, tomando en consideración que $\binom{m}{k} = 0$ cuando $m < k$, se tiene $f_X(x) = \frac{\binom{r}{x}\binom{s}{n-x}}{\binom{r+s}{n}}$ para cualquier $x \in \{0, \dots, n\}$.

El nombre hipergeométrica obedece a las siguientes consideraciones:

Para simplificar, supongamos que $n < \text{mín}\{r, s\}$. Se tiene entonces, $\binom{s}{n-k}\binom{s-n+k}{k} = \binom{s}{n}\binom{n}{k}$, para cualquier $k \in \{\text{máx}\{0, n-s\}, \dots, \text{mín}\{n, r\}\}$, así que:

$$\frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}} = \frac{\binom{r}{k}}{\binom{r+s}{n}} \frac{\binom{s}{n}\binom{n}{k}}{\binom{s-n+k}{k}} = \frac{\binom{s}{n}}{\binom{r+s}{n}} \frac{\binom{n}{k}\binom{r}{k}}{\binom{s-n+k}{k}}$$

Por otra parte, se puede demostrar que si a, b son números enteros cualesquiera y c un número entero positivo, entonces la serie $\sum_{k=0}^{\infty} \frac{\binom{-a}{k}\binom{-b}{k}}{\binom{c+k-1}{k}} x^k$ converge absolutamente para cualquier x tal que $|x| < 1$. Además, si $c > b > 0$, se tiene la siguiente representación integral:

$$\sum_{k=0}^{\infty} \frac{\binom{-a}{k}\binom{-b}{k}}{\binom{c+k-1}{k}} x^k = (c-b)\binom{c-1}{b-1} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tx)^{-a} dt$$

El término $(1-tx)^{-a}$, el cual, para $a = 1$, se desarrolla como una serie geométrica, motiva que a la serie del lado izquierdo de la última igualdad se le llame serie hipergeométrica.

Finalmente, obsérvese que los términos $\frac{\binom{n}{k}\binom{r}{k}}{\binom{s-n+k}{k}}$ corresponden a los de una serie hipergeométrica en donde $a = -n$, $b = -r$ y $c = s - n + 1$.

La distribución hipergeométrica surgió prácticamente desde los inicios del Cálculo de Probabilidades, al considerarse el experimento aleatorio consistente en seleccionar, con reemplazo, bolas de una urna, la cual contiene bolas de dos colores, y buscarse la probabilidad de obtener cierto número de bolas de uno de los colores. Un problema de este tipo se encuentra formulado en el libro de C. Huygens:

A tiene doce fichas, 4 blancas y 8 negras, y apuesta con B a que al tomar 7 fichas al azar, 3 de ellas serán blancas. ¿Cuál es la probabilidad de que A gane?

De manera general, la distribución hipergeométrica se presenta al considerar un muestreo sin reemplazo de la siguiente manera: Supongamos que se tiene una población formada por dos tipos de elementos, los cuales llamaremos de tipo I y II, de tal manera que hay r elementos de tipo I y s de tipo II. Si se toma una muestra sin reemplazo de tamaño $n \leq r+s$ de esa población y llamamos X al número de elementos de tipo I que se obtienen en la muestra, entonces X tiene distribución hipergeométrica con parámetros r, s y n. En efecto, el número total de posibles muestras que pueden obtenerse es igual a $\binom{r+s}{n}$ y todas ellas son equiprobables. De éstas, para $k \in \{\text{máx}\{0, n-s\}, \ldots, \text{mín}\{n, r\}\}$, el total de muestras en las cuales se obtienen k objetos de tipo I y $n-k$ objetos de tipo II es igual a $\binom{r}{k}\binom{s}{n-k}$, de lo cual se sigue el resultado (ver ejemplo 3.10).

PROPOSICIÓN 7.21. *Los términos* $t_k = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}$, *de una distribución hipergeométrica, crecen con* k *hasta alcanzar su máximo valor cuando* $\frac{(n+1)(r+1)}{r+s+2} - 1 \leq k \leq \frac{(n+1)(r+1)}{r+s+2}$, *después de lo cual decrecen con* k.

Demostración

Sea $m = \text{máx}\{0, n-s\}$, $M = \text{mín}\{n, r\}$ y, para $k \in \{m, \ldots, M\}$, $t_k = \frac{\binom{r}{x}\binom{s}{n-x}}{\binom{r+s}{n}}$. Se tiene entonces, para $k \in \{m+1, \ldots, M\}$:

$\frac{t_{k-1}}{t_k} = \frac{k(s-n+k)}{(n-k+1)(r-k+1)}$

De manera que $t_{k-1} \leq t_k$ si y sólo si $k(s-n+k) \leq (n-k+1)(r-k+1)$, es decir, $k \leq \frac{(n+1)(r+1)}{r+s+2}$.

Para $k \in \{m, m+1, \ldots, M-1\}$, se tiene:

$\frac{t_k}{t_{k+1}} = \frac{(k+1)(s-n+k+1)}{(n-k)(r-k)}$

De manera que $t_k \geq t_{k+1}$ si y sólo si $(k+1)(s-n+k+1) \geq (n-k)(r-k)$, es decir, $k \geq \frac{(n+1)(r+1)}{r+s+2} - 1$.

∎

A continuación se presentan las gráficas de algunas funciones de densidad tipo hipergeométrica. Con el objeto de visualizar más fácilmente la forma que adquieren, se han trazado continuas, graficando la función:

$f(x) = \frac{\Gamma(r+1)}{\Gamma(x+1)\Gamma(r-x+1)} \frac{\Gamma(s+1)}{\Gamma(n-x+1)\Gamma(s-n+x+1)} \frac{\Gamma(n+1)\Gamma(r+s-n+1)}{\Gamma(r+s+1)}$

definida en el intervalo $[0, n]$. Evidentemente, los valores de la función de densidad corresponden únicamente a los valores enteros comprendidos entre 0 y n inclusive.

$r = 10, s = 90, n = 20$ $\quad$ $r = 90, s = 10, n = 20$ $\quad$ $r = 80, s = 120, n = 20$

EJEMPLO 7.22. *En un lago se capturan al azar* 1000 *peces, los cuales se marcan y se devuelven al lago. Después de algún tiempo, se capturan al azar otros* 1000 *peces y entre ellos se encuentran* 100 *peces de los marcados. Con este dato, se quiere estimar el número de peces que hay en el lago. Para esto, denotemos por $p(k,n)$ a la probabilidad de obtener k peces marcados en la muestra cuando hay n peces en el lago, se tiene entonces:*

$p(k,n) = \frac{\binom{1000}{k}\binom{n-1000}{1000-k}}{\binom{n}{1000}}$

$\frac{p(k,n)}{p(k,n-1)} = \frac{(n-1000)^2}{n(n-2000+k)} = 1 + \frac{1000000-nk}{n(n-2000+k)}$

De manera que, fijando k, $p(k,n)$ crece con n hasta alcanzar su valor máximo cuando $\frac{1000000}{k} - 1 \leq n \leq \frac{1000000}{k}$, después de lo cual decrece con n.

Un método de estimación puede consistir entonces en tomar como valor de n uno que haga máximo el valor de $p(k,n)$. En nuestro caso tenemos $k = 100$, así que un valor de n que hace máximo el valor de $p(k,n)$ es $n = 10,000$.

▲

Como ya lo hemos hecho notar con anterioridad, para poblaciones grandes, prácticamente no hay diferencia entre el muestreo con reemplazo y el muestreo sin reemplazo. Esto significa que, en ese caso, una distribución binomial y una hipergeométrica serán muy parecidas (ver sección 3.3). A continuación se prueba este resultado.

TEOREMA 7.23. *Supongamos que, para cada $r, s, n \in \mathbb{N}$, $X_n^{(r,s)}$ es una variable aleatoria con distribución hipergeométrica de parámetros r, s y n, entonces, para cualquier $k \in \{0, 1, \ldots, n\}$, se tiene:*

$$\lim_{r,s \rightsquigarrow \infty} \frac{P\left[X_n^{(r,s)}=k\right]}{\binom{n}{k}\left(\frac{r}{r+s}\right)^k\left(\frac{s}{r+s}\right)^{n-k}} = 1$$

Demostración

Suponiendo $n < \text{mín}\{r,s\}$, se tiene, para $k \in \{0,1,2,\ldots,n\}$:

$\binom{r}{k} = \frac{r(r-1)\cdots(r-k+1)}{k!} = \frac{r(r-1)\cdots(r-k+1)}{r^k}\frac{r^k}{k!} < \frac{r^k}{k!}$

$\binom{s}{n-k} = \frac{s(s-1)\cdots(s-n+k+1)}{(n-k)} = \frac{s(s-1)\cdots(s-n+k+1)}{s^{n-k}}\frac{s^{n-k}}{(n-k)!} < \frac{s^{n-k}}{(n-k)!}$

$\binom{r+s}{n} = \frac{(r+s)(r+s-1)\cdots(r+s-n+1)}{n!} < \frac{(r+s)^n}{n!}$

$\binom{r}{k} = \frac{r(r-1)\cdots(r-k+1)}{k!} = \frac{r(r-1)\cdots(r-k+1)}{r^k}\frac{r^k}{k!} > \frac{(r-k)^k}{r^k}\frac{r^k}{k!} = \left(1-\frac{k}{r}\right)^k\frac{r^k}{k!}$

$\binom{s}{n-k} = \frac{s(s-1)\cdots(s-n+k+1)}{(n-k)} = \frac{s(s-1)\cdots(s-n+k+1)}{s^{n-k}}\frac{s^{n-k}}{(n-k)!} > \frac{(s-n+k)^k}{s^{n-k}}\frac{s^{n-k}}{(n-k)!} = \left(1-\frac{n-k}{s}\right)^k\frac{s^{n-k}}{(n-k)!}$

$\binom{r+s}{n} = \frac{(r+s)(r+s-1)\cdots(r+s-n+1)}{n!} = \frac{(r+s)(r+s-1)\cdots(r+s-n+1)}{(r+s)^n}\frac{(r+s)^n}{n!} > \frac{(r+s-n)^n}{(r+s)^n}\frac{(r+s)^n}{n!} = \left(1-\frac{n}{r+s}\right)^n\frac{(r+s)^n}{n!}$

Así que:

$f_{X_n^{(r,s)}}(k) = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}} < \frac{r^k}{k!}\frac{s^{n-k}}{(n-k)!}\frac{n!}{(r+s)^n}\left(1-\frac{n}{r+s}\right)^{-n}$

$= \left(1-\frac{n}{r+s}\right)^{-n}\frac{n!}{k!(n-k)!}\frac{r^k}{(r+s)^k}\frac{s^{n-k}}{(r+s)^{n-k}} = \left(1-\frac{n}{r+s}\right)^{-n}\binom{n}{k}\left(\frac{r}{r+s}\right)^k\left(\frac{s}{r+s}\right)^{n-k}$

$f_{X_n^{(r,s)}}(k) = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}} > \left(1-\frac{k}{r}\right)^k\frac{r^k}{k!}\left(1-\frac{n-k}{s}\right)^k\frac{s^{n-k}}{(n-k)!}\frac{n!}{(r+s)^n}$

$= \left(1-\frac{k}{r}\right)^k\left(1-\frac{n-k}{s}\right)^{n-k}\frac{n!}{k!(n-k)!}\frac{r^k}{(r+s)^k}\frac{s^{n-k}}{(r+s)^{n-k}}$

$= \left(1-\frac{k}{r}\right)^k\left(1-\frac{n-k}{s}\right)^{n-k}\binom{n}{k}\left(\frac{r}{r+s}\right)^k\left(\frac{s}{r+s}\right)^{n-k}$

Es decir:

$\left(1-\frac{k}{r}\right)^k\left(1-\frac{n-k}{s}\right)^{n-k}\binom{n}{k}\left(\frac{r}{r+s}\right)^k\left(\frac{s}{r+s}\right)^{n-k} < f_{X_n^{(r,s)}}(k) < \left(1-\frac{n}{r+s}\right)^{-n}\binom{n}{k}\left(\frac{r}{r+s}\right)^k\left(\frac{s}{r+s}\right)^{n-k}$

Por lo tanto:

$\text{lím}_{r,s\rightsquigarrow\infty}\frac{P\left[X_n^{(r,s)}=k\right]}{\binom{n}{k}\left(\frac{r}{r+s}\right)^k\left(\frac{s}{r+s}\right)^{n-k}} = 1$

■

Con el objeto de comparar los valores exactos de una función de densidad hipergeométrica con sus valores aproximados por medio de una densidad binomial, a continuación se muestran las gráficas que se obtienen con ambos valores. Con el objeto de apreciar mejor la aproximación, las gráfica de los valores exactos y aproximados de la densidad hipergeométrica $f_X(x)$ se han trazado continuas, expresándolos como $\frac{\Gamma(r+1)}{\Gamma(x+1)\Gamma(r-x+1)}\frac{\Gamma(s+1)}{\Gamma(n-x+1)\Gamma(s-n+x+1)}\frac{\Gamma(n+1)\Gamma(r+r-n+1)}{\Gamma(r+s+1)}$ y $\frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}p^x(1-p)^{n-x}$, respectivamente. La línea sólida corresponde a los valores exactos, mientras que la línea punteada corresponde a los valores aproximados.

$r=4, s=6, n=5$ $\qquad$ $r=90, s=10, n=20$ $\qquad$ $r=80, s=120, n=20$

Ejemplo 7.24. *Supongamos que en un supermercado se venden* 400 *artículos,* 6 *de los cuales no tienen marcada la clave de su precio. Si un cliente compra* 10 *artículos, estime la probabilidad de que entre ellos haya por lo menos uno que no tenga marcada la clave de su precio.*

Solución

Sea X el número de artículos, de los 10 *que compra el cliente, que no tienen marcada la clave de su precio. X tiene entonces distribución hipergeométrica con parámetros $r=6$, $s=394$ y $n=10$, la cual se puede aproximar mediante una distribución binomial con parámetros $n=10$ y $p=\frac{r}{r+s}=0.015$. A su vez ésta se puede aproximar mediante una distribución Poisson con parámetro $\lambda=np=(10)(0.015)=0.15$, así que:*

$$P[X\geq 1]=1-P[X=0]\approx 1-e^{-\lambda}=1-e^{-0.15}=0.13929$$

Aproximando la distribución de X mediante una binomial con parámetros $n=10$ y $p=0.015$, se obtiene:

$$P[X\geq 1]=1-P[X=0]=1-(1-p)^{10}=1-(0.985)^{10}=0.14027$$

Utilizando la distribución exacta (hipergeométrica), se obtiene:

$$P[X\geq 1]=1-P[X=0]=1-\frac{\binom{6}{0}\binom{394}{10}}{\binom{400}{10}}=0.14177$$

Ejemplo 7.25. *Supongamos que el* 5 % *de las personas que manejan tienen su licencia de manejo vencida. Utilice la aproximación de Poisson para estimar la probabilidad de que a lo más* 2 *de* 50 *personas que manejan, seleccionadas al azar, tengan su licencia de manejo vencida.*

Solución

Sea X el número de personas, del grupo de 50, que tienen su licencia de manejo vencida. X tiene entonces, aproximadamente, distribución Poisson con parámetro $\lambda = (50)(0.05) = 2.5$, así que:

$P[X \leq 2] \approx \sum_{k=0}^{2} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}(1 + \lambda + \frac{\lambda^2}{2}) = \frac{53}{8}e^{-5/2} = 0.543813$

EJEMPLO 7.26. *Una máquina produce artículos defectuosos y no defectuosos de tal manera que, en promedio, se producen 20 defectuosos en cada 1000 artículos. ¿Cuál es la probabilidad de que una caja con 100 artículos, seleccionados al azar de una población grande, no contenga artículos defectuosos?*

Solución

Sea N el total de artículos en la población y llamemos X al número de artículos defectuosos en la caja con 100 artículos. Entonces:

$P[X = k] = \frac{\binom{0.02N}{k}\binom{0.98N}{100-k}}{\binom{N}{100}} \approx \binom{100}{k}(0.02)^k(0.98)^{100-k} \approx \frac{2^k e^{-2}}{k!}$

Así que:

$P[X = 0] \approx e^{-2} = 0.1353$

7.6. Otras distribuciones

7.6.1. Distribuciones truncadas. En ocasiones, una distribución discreta se presenta de manera truncada, lo cual significa que algunos de sus posibles valores no son tomados en cuenta. Un caso frecuente se presenta cuando el valor 0 no se considera. Por ejemplo, al estudiar la atención diaria a usuarios de una determinada institución, los días laborables en los cuales el número de usuarios atendidos es cero podrían no ser considerados por corresponder a días en los cuales la institución no atendió a usuarios por causas que le fueron ajenas.

De esta manera, por cada una de las distribuciones estudiadas anteriormente se puede definir la correspondiente distribución truncada. Por ejemplo, una variable aleatoria X con distribución binomial truncada, mediante la eliminación del valor 0, tiene una función de densidad dada por:

$$f_X(x) = \begin{cases} c\binom{n}{x}p^x(1-p)^{n-x} & \text{si } x \in \{1, \dots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde c es una constante.

El valor de c se puede determinar fácilmente utilizando el hecho de que f_X es una función de densidad.

Un caso interesante es el de una variable aleatoria X con distribución binomial negativa, de parámetros $p \in (0,1]$ y $r \in (0,\infty)$, truncada, mediante la eliminación del valor 0. En este caso la función de densidad de X está dada por:

$$f_X(x) = \begin{cases} c\frac{\Gamma(r+x)}{x!\Gamma(r)}p^r(1-p)^x & \text{si } x \in \{1,2,\ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde c es una constante.

Como $\sum_{k=1}^{\infty} \frac{\Gamma(k+r)}{k!\Gamma(r)}p^r(1-p)^k = 1 - p^r$, se obtiene que $c = \frac{1}{1-p^r}$, de manera que se tiene:

$$f_X(x) = \begin{cases} \frac{\Gamma(r+x)}{x!\Gamma(r)}\frac{p^r}{1-p^r}(1-p)^x & \text{si } x \in \{1,2,\ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

El caso $r = 0$ no está incluido aquí. Sin embargo, la distribución para este caso puede obtenerse tomando límites. En efecto, para $x \in \{1,2,\ldots\}$, se tiene:

$\text{lím}_{r\rightsquigarrow 0} f_X(x) = \text{lím}_{r\rightsquigarrow 0} \frac{\Gamma(r+x)}{x!\Gamma(r)}\frac{p^r}{1-p^r}(1-p)^x = \text{lím}_{r\rightsquigarrow 0} \frac{\Gamma(r+x)}{x!\Gamma(r+1)}\frac{r}{1-p^r}p^r(1-p)^x$

$= \frac{\Gamma(x)}{x!}\frac{1}{-\ln p}(1-p)^x = -\frac{1}{\ln(1-q)}\frac{q^x}{x}$

Se puede ver que lo que se obtiene como límite es una función de densidad (ver ejercicio 7.30), la cual es conocida como densidad serie logarítmica por el hecho de que los términos $\frac{q^k}{k}$ corresponden a los del desarrollo, en serie de Taylor alrededor de 0, de la función $\ln(1-q)$.

7.6.2. Distribución uniforme discreta.

Definición 7.27 (**Distribución uniforme discreta**). *Si A es un conjunto finito de números reales, se dice que una variable aleatoria X tiene distribución uniforme discreta en el conjunto A si su función de densidad está dada por la fórmula:*

$$f_X(x) = \begin{cases} \frac{1}{N} & \text{si } x \in A \\ 0 & \text{en otro caso} \end{cases}$$

en donde N es el número de elementos de A.

Como puede verse inmediatamente, una distribución uniforme discreta surge al considerar una variable X definida como el número que se obtiene al seleccionar al azar un elemento de un conjunto finito de números reales. Es decir, la distribución uniforme es una expresión de la elección al azar, la cual a su vez es sinónimo de equiprobabilidad.

Por su simplicidad, la distribución uniforme puede ser utilizada para ilustrar algunos métodos para encontrar probabilidades de eventos que dependen de varias variables aleatorias. Esto se muestra en los siguientes dos ejemplos:

EJEMPLO 7.28. *Sean X y Y dos variables aleatorias independientes, X con distribución uniforme en el conjunto $\{0, \ldots, n\}$, Y con distribución binomial de parámetros n y p. a) Encuentre $P[X < Y]$ y $P[X > Y]$. b) Determine los valores de n y p para los cuales $P[X > Y] > P[X < Y]$.*

Solución

$P[X < Y] = \sum_{y=0}^{n} P[X < Y, Y = y] = \sum_{y=0}^{n} P[X < y, Y = y]$

$= \sum_{y=0}^{n} P[X < y] P[Y = y] = \sum_{y=0}^{n} \frac{y}{n+1} P[Y = y] = \frac{1}{n+1} \sum_{y=0}^{n} y P[Y = y] = \frac{np}{n+1}$

$P[X > Y] = \sum_{y=0}^{n} P[X > Y, Y = y] = \sum_{y=0}^{n} P[X > y, Y = y]$

$= \sum_{y=0}^{n} P[X > y] P[Y = y] = \sum_{y=0}^{n} \frac{n-y}{n+1} P[Y = y]$

$= \frac{n}{n+1} \sum_{y=0}^{n} P[Y = y] - \frac{1}{n+1} \sum_{y=0}^{n} y P[Y = y] = \frac{n}{n+1} - \frac{np}{n+1} = \frac{n(1-p)}{n+1}$

Así que, $P[X > Y] > P[X < Y]$ cuando $\frac{n(1-p)}{n+1} > \frac{np}{n+1}$, es decir, $p < \frac{1}{2}$.

EJEMPLO 7.29. *Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el conjunto $\{1, \ldots, N\}$. Encuentre la función de densidad de $Z = \text{mín}(X, Y)$.*

Solución

Si $z \in \{1, \ldots, N\}$, se tiene:

$P[Z = z] = P[\text{mín}(X, Y) = z]$

$= P[X = z, Y > z] + P[X > z, Y = z] + P[X = z, Y = z]$

$= \frac{1}{N} \frac{N-z}{N} + \frac{N-z}{N} \frac{1}{N} + \frac{1}{N} \frac{1}{N} = \frac{2(N-z)+1}{N^2}$

Por lo tanto:

$$f_Z(z) = \begin{cases} \frac{2(N-z)+1}{N^2} & \text{si } z \in \{1, \ldots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

7.7. Caminatas aleatorias

Consideremos una partícula que se mueve a saltos sobre la recta real de tal manera que, comenzando en el origen, en cada intervalo de tiempo h, la partícula salta una distancia d hacia la derecha con probabilidad $\frac{1}{2}$ o una distancia d hacia la izquierda, también con probabilidad $\frac{1}{2}$. Llamemos S_n a la posición de la partícula en el tiempo nh. La familia de variables aleatorias S_n es uno de los ejemplos más simples de lo que se conoce con el nombre de caminata aleatoria.

DEFINICIÓN 7.30 (**Caminata aleatoria**). *Dados un número real z y una sucesión de variables aleatorias, $Z_1, Z_2, \ldots$, idénticamente distribuidas, y tales que cualquier colección finita de ellas forma una familia de variables aleatorias independientes, la sucesión de variables aleatorias $S_n = z + \sum_{k=1}^{n} Z_k$ es llamada una caminata aleatoria.*

Las caminatas aleatorias juegan un papel muy importante en la Teoría de la Probabilidad moderna pues modelan, por lo menos como una aproximación, muchos fenómenos aleatorios de interés.

En esta parte, para el caso particular de la caminata aleatoria introducida al inicio de esta sección, nos limitaremos a encontrar la distribución de tres variables aleatorias de interés, a saber, S_n misma, el tiempo que lleva a la partícula regresar al origen y el tiempo que lleva a la partícula alcanzar una posición positiva. De paso abordaremos el problema de determinar el número de retornos al origen que se realizan, cuando se permite que la partícula se mueva indefinidamente.

7.7.1. Distribución de la posición en el n-simo paso. Para cada $n \in \mathbb{N}$, definamos las variables aleatorias X_n y Y_n como el número de saltos que da la partícula hacia la derecha y hacia la izquierda, respectivamente, en el tiempo nh.

Si consideramos cada salto de la partícula como un ensayo de Bernoulli, en donde un movimiento hacia la derecha representa un éxito, X_n resulta ser entonces igual al número de éxitos que se tienen en n ensayos, así que su distribución es binomial con parámetros n y $p = \frac{1}{2}$, es decir:

$$P[X_n = x] = \begin{cases} \binom{n}{x}\frac{1}{2^n} & \text{si } x \in \{0, \ldots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

Ahora bien, se tiene:

$X_n + Y_n = n$ y $d(X_n - Y_n) = S_n$, así que, $S_n = d(2X_n - n)$.

Por lo tanto, se tiene el siguiente resultado:

PROPOSICIÓN 7.31. $P[S_n = dr] = P\left[X_n = \frac{n+r}{2}\right]$

$$= \begin{cases} \binom{n}{\frac{n+r}{2}} \frac{1}{2^n} & \textit{si } r \in \{-n, -n+2, \ldots, n-2, n\} \\ 0 & \textit{en otro caso} \end{cases}$$

Este resultado se puede interpretar geométricamente de la siguiente manera:

El movimiento de la partícula puede representarse en un plano cartesiano tomando el tiempo sobre el eje horizontal y la posición en cada tiempo sobre el eje vertical. Además, para simplificar, podemos tomar h como unidad en el eje horizontal y d en el eje vertical. De esta forma, los posibles movimientos de la partícula son similares a los de la siguiente figura:

A cada sucesión de puntos $(1, S_1), (2, S_2), \ldots, (n, S_n)$ la llamaremos una textbftrayectoria hasta el paso n.

Ahora bien, es claro que el número total de posibles trayectorias hasta el paso n es igual a 2^n y todas ellas tienen la misma probabilidad de ocurrencia. La relación $P[S_n = dr] = \binom{n}{\frac{n+r}{2}} \frac{1}{2^n}$ expresa entonces el siguiente resultado:

PROPOSICIÓN 7.32. *Si* $r \in \{-n, -n+2, \ldots, n-2, n\}$, *entonces el número de posibles trayectorias que conducen a la partícula a la posición* (n, r) *es igual a* $\binom{n}{\frac{n+r}{2}}$.

7.7.2. Retornos al origen.

DEFINICIÓN 7.33 (**Retornos al origen**). *Diremos que hay un retorno al origen en el paso* n *si* $S_n = 0$.

Obsérvese que si hay un retorno al origen en el paso n, entonces n es un número par. Ahora bien, utilizando la imagen geométrica descrita en la subsección 7.7.1, para que no se presente ningún retorno al origen hasta el paso n se requiere que, a partir del paso 1 y hasta el paso n, la partícula se mantenga ya sea todo el tiempo por arriba del eje horizontal, o bien todo el tiempo por debajo de éste. Como todas las trayectorias hasta el paso n son igualmente probables, para calcular la probabilidad de que ocurra alguna de las dos situaciones anteriores basta con contar el total de trayectorias que producen su respectiva ocurrencia. La siguiente proposición establece el resultado de tal conteo.

PROPOSICIÓN 7.34. *Sea (n, r) un punto en el primer cuadrante que sea un posible punto de paso de la caminata aleatoria, entonces el total de posibles trayectorias tales que $S_1 > 0$, $S_2 > 0$, $\ldots$, $S_{n-1} > 0$ y que conducen a la partícula a la posición (n, r), es igual a* $\binom{n-1}{\frac{n+r}{2}-1} - \binom{n-1}{\frac{n+r}{2r>0}} = \frac{r}{n}\binom{n}{\frac{n+r}{2}}$.

Demostración

Como (n, r) es un posible punto de paso en el cuadrante positivo, se tiene $r > 0$ y y $r \in \{n, n-2, \ldots\}$. Ahora bien, llamemos $K_{n,r}^+$ al total de posibles trayectorias tales que $S_1 > 0$, $S_2 > 0$, $\ldots$, $S_{n-1} > 0$ y que conducen a la partícula a la posición (n, r); $K_{n,r}^+$ es entonces igual al total de posibles trayectorias que van del punto $(1, 1)$ al punto (n, r) y que no tocan el eje horizontal. Pero hay una correspondencia uno a uno entre el conjunto de posibles trayectorias que van del punto $(1, 1)$ al punto (n, r) y que tocan el eje horizontal con el conjunto total de posibles trayectorias que van del punto $(1, -1)$ al punto (n, r), pues dada una posible trayectoria τ que va del punto $(1, 1)$ al punto (n, r) y que toca el eje horizontal en el punto $(t, 0)$, se le puede asociar la trayectoria τ' que comienza en $(1, -1)$ y conduce al punto (n, r), pasando por $(t, 0)$, definida como el reflejo de la trayectoria τ en el eje horizontal hasta el punto $(t, 0)$ y que coincide con τ desde el punto $(t, 0)$ hasta el punto (n, r), como se ilustra en la siguiente figura:

Con base en lo anterior, $K_{n,r}^+$ es igual al total de posibles trayectorias que van del punto $(1, 1)$ al punto (n, r) menos el total de posibles trayectorias que van del punto $(1, -1)$ al punto (n, r). Pero hay tantas trayectorias que van del punto $(1, 1)$ (resp.

$(1,-1)$) al punto (n,r) como trayectorias que van del origen al punto $(n-1,r-1)$ (resp. $(n-1,r+1)$), de manera que, por la proposición 7.32, se tiene, para $r \neq n$:

$$K_{n,r}^{+} = \binom{n-1}{\frac{n+r-2}{2}} - \binom{n-1}{\frac{n+r}{2}} = \binom{n-1}{\frac{n+r}{2}-1} - \binom{n-1}{\frac{n+r}{2}} = \frac{n+r}{2n}\binom{n}{\frac{n+r}{2}} - \frac{n-r}{2n}\binom{n}{\frac{n+r}{2}} = \frac{r}{n}\binom{n}{\frac{n+r}{2}}$$

Además, como $K_{n,n}^{+} = 1$, el resultado también se cumple para $r = n$.

■

PROPOSICIÓN 7.35. *La probabilidad de que no ocurra ningún retorno al origen hasta el paso $2m$, inclusive, es igual a $\binom{2m}{m}\frac{1}{2^{2m}}$.*

Demostración

Sea p_m la probabilidad de que no ocurra ningún retorno al origen hasta el paso $2m$, entonces, utilizando la proposición 7.34, se tiene:

$$p_m = P\left[S_1 \neq 0, S_2 \neq 0, \ldots, S_{2m} \neq 0\right]$$

$$= P\left[S_1 > 0, S_2 > 0, \ldots, S_{2m} > 0\right] + P\left[S_1 < 0, S_2 < 0, \ldots, S_{2m} < 0\right]$$

$$= 2P\left[S_1 > 0, S_2 > 0, \ldots, S_{2m} > 0\right]$$

$$= 2\sum_{x=1}^{m} P\left[S_1 > 0, S_2 > 0, \ldots, S_{2m-1} > 0, S_{2m} = 2x\right]$$

$$= \frac{2}{2^{2m}}\sum_{x=1}^{m}\left[\binom{2m-1}{m+x-1} - \binom{2m-1}{m+x}\right] = \frac{2}{2^{2m}}\left[\binom{2m-1}{m} - \binom{2m-1}{2m}\right]$$

$$= \frac{2}{2^{2m}}\binom{2m-1}{m} = \frac{1}{2^{2m}}\binom{2m}{m}$$

■

Obsérvese que si definimos $p_m = \binom{2m}{m}\frac{1}{2^{2m}}$, entonces, utilizando el corolario 6.48, se tiene:

$$\text{lím}_{m\rightsquigarrow\infty}\, p_m = \text{lím}_{m\rightsquigarrow\infty}\binom{2m}{m}\frac{1}{2^{2m}}$$

$$= \text{lím}_{m\rightsquigarrow\infty}\frac{(2m)!}{m!m!}\frac{1}{2^{2m}} = \text{lím}_{m\rightsquigarrow\infty}\frac{(2m)!\sqrt{m}}{2^{2m}(m!)^2}\frac{1}{\sqrt{m}}$$

$$= \text{lím}_{m\rightsquigarrow\infty}\frac{(2m)!\sqrt{m}}{2^{2m}(m!)^2}\,\text{lím}_{m\rightsquigarrow\infty}\frac{1}{\sqrt{m}} = \frac{1}{\sqrt{\pi}}\,\text{lím}_{m\rightsquigarrow\infty}\frac{1}{\sqrt{m}} = 0$$

PROPOSICIÓN 7.36. *Para cualquier $n \in \mathbb{N}$, la probabilidad de que haya por lo menos n retornos al origen es igual a 1.*

Demostración

Sea R_n el paso en que ocurre el n-simo retorno al origen y definamos los eventos:

B_n: Ocurren menos de n retornos al origen.

A_0: No ocurre ningún retorno al origen.

A_0^m: No ocurre retorno al origen en ninguno de los primeros $2m$ pasos.

A_n: No ocurre ningún retorno al origen después del n-simo retorno al origen.

A_n^m: No ocurre retorno al origen en ninguno de los $2m$ pasos que siguen al n-simo retorno al origen.

De acuerdo con la proposición 7.35, se tiene $P[A_0] \leq P[A_0^m] = p_m$ para cualquier m. Por lo tanto, $P[A_0] = 0$.

Supongamos ahora que $P[A_0] = P[A_1] = \cdots = P[A_k] = 0$, entonces:

$P[B_{k+1}] = P[A_0 \cup A_1 \cup \cdots \cup A_k] \leq \sum_{j=1}^{k} P[A_j] = 0$

Pero, si ocurre B_{k+1}^c, entonces R_{k+1} es finito. Por lo tanto:

$P[R_{k+1} < \infty] \geq P\left[B_{k+1}^c\right] = 1$

Así que, para cualquier m, se tiene:

$P[A_{k+1}] = P[A_{k+1}, R_{k+1} < \infty] = \sum_{j=1}^{\infty} P[A_{k+1}, R_{k+1} = 2j]$

$\leq \sum_{j=1}^{\infty} P\left[A_{k+1}^m, R_{k+1} = 2j\right] = \sum_{j=1}^{\infty} P\left[A_{k+1}^m \mid R_{k+1} = 2j\right] P[R_{k+1} = 2j]$

$= \sum_{j=1}^{\infty} P[A_0^m] P[R_{k+1} = 2j] = P[A_0^m] \sum_{j=1}^{\infty} P[R_{k+1} = 2j]$

$= P[A_0^m] P[R_{k+1} < \infty] = P[A_0^m] = p_m$

Se puede concluir entonces que $P[A_{k+1}] \leq \text{lím}_{m \rightsquigarrow \infty} p_m = 0$.

Así que, por el principio de inducción, $P[A_n] = 0$ para cualquier $n \in \{0, 1, \ldots\}$.

Finalmente, para cualquier $n \in \mathbb{N}$, se tiene:

$P[B_n] = P[A_0 \cup A_1 \cup \cdots \cup A_{n-1}] \leq \sum_{j=1}^{n-1} P[A_j] = 0$

Así que, $P[B_n^c] = 1$

■

Corolario 7.37. *La probabilidad de que haya una infinidad de retornos al origen es igual a* 1.

Demostración

Definamos los eventos:

B_n: Ocurren menos de n retornos al origen.

B: Se produce únicamente un número finito de retornos al origen.

Obsérvese que $B_1 \subset B_2 \subset \cdots$ y

$B = \bigcup_{m=1}^{\infty} B_m = B_1 \cup (B_2 - B_1) \cup (B_3 - B_2) \cup \cdots$

Así que:

$P(B) = P(B_1) + P(B_2 - B_1) + P(B_3 - B_2) + \cdots = 0$

■

7.7.3. Distribución del tiempo del primer retorno al origen.

Sea R el paso en que ocurre el primer retorno al origen. Por la proposición 7.36, $P[R < \infty] = 1$. Además, utilizando la proposición 7.35, se tiene:

$P[R = 2m] = P[S_1 \neq 0, S_1 \neq 0, \ldots, S_{2m-1} \neq 0, S_{2m} = 0]$

$= P[S_1 \neq 0, S_1 \neq 0, \ldots, S_{2m-2} \neq 0] - P[S_1 \neq 0, S_1 \neq 0, \ldots, S_{2m} \neq 0]$

$= \frac{1}{2^{2m-2}}\binom{2m-2}{m-1} - \frac{1}{2^{2m}}\binom{2m}{m} = \frac{1}{2^{2m-2}}\frac{m^2}{(2m-1)(2m)}\binom{2m}{m} - \frac{1}{2^{2m}}\binom{2m}{m}$

$= \frac{2m}{2m-1}\frac{1}{2^{2m}}\binom{2m}{m} - \frac{1}{2^{2m}}\binom{2m}{m} = \left[\frac{2m}{2m-1} - 1\right]\frac{1}{2^{2m}}\binom{2m}{m} = \frac{1}{2m-1}\frac{1}{2^{2m}}\binom{2m}{m}$

7.7.4. Primer paso por un valor positivo.

PROPOSICIÓN 7.38. *La probabilidad de que no ocurra ningún paso por un valor positivo en ninguno de los primeros $2m$ pasos, inclusive, es igual a $\binom{2m}{m}\frac{1}{2^{2m}}$.*

Demostración

$P[S_1 < 0, S_2 < 0, \ldots, S_{2m} < 0]$

$= P[X_1 = -1, X_2 \leq 0, X_2 + X_3 \leq 0, \ldots, X_2 + X_3 + \cdots + X_{2m} \leq 0]$

$= P[X_1 = -1]\, P[X_2 \leq 0, X_2 + X_3 \leq 0, \ldots, X_2 + X_3 + \cdots + X_{2m} \leq 0]$

$= \frac{1}{2}P[X_1 \leq 0, X_1 + X_2 \leq 0, \ldots, X_1 + X_2 + \cdots + X_{2m-1} \leq 0]$

$= \frac{1}{2}P[S_1 \leq 0, S_2 \leq 0, \ldots, S_{2m-1} \leq 0] = \frac{1}{2}P[S_1 \leq 0, S_2 \leq 0, \ldots, S_{2m} \leq 0]$

Así que:

$P[S_1 \leq 0, S_2 \leq 0, \ldots, S_{2m} \leq 0] = 2P[S_1 < 0, S_2 < 0, \ldots, S_{2m} < 0]$

Además:

$P[S_1 < 0, S_2 < 0, \ldots, S_{2m} < 0] = \frac{1}{2} P[S_1 \neq 0, S_2 \neq 0, \ldots, S_{2m} \neq 0]$

Por lo tanto:

$P[S_1 \leq 0, S_2 \leq 0, \ldots, S_{2m} \leq 0] = P[S_1 \neq 0, S_2 \neq 0, \ldots, S_{2m} \neq 0] = \binom{2m}{m} \frac{1}{2^{2m}}$

■

Proposición 7.39. *La probabilidad de que haya por lo menos un paso por un valor positivo es igual a* 1.

Demostración

Definamos los eventos:

A_0: No ocurre ningún paso por un valor positivo.

A_0^m: No ocurre ningún paso por un valor positivo en ninguno de los primeros $2m$ pasos.

Se tiene, $P[A_0] \leq P[A_0^m] = p_m$ para cualquier $m \in \mathbb{N}$, por lo tanto:

$P[A_0] \leq \lim_{m \rightsquigarrow \infty} p_m = 0$

■

Corolario 7.40. *Para cualquier $n \in \mathbb{Z}$, la probabilidad de que haya una infinidad de pasos por la posición $z = n$ es igual a* 1.

Demostración

Para cada $n \in \mathbb{N}$, Sea Z_n el paso en el cual la partícula alcanza el valor $z = n$ por primera vez, entonces, por la proposición 7.39, $P[Z_1 < \infty] = 1$.

Supongamos ahora que $P[Z_k < \infty] = 1$ para alguna $k \in \mathbb{N}$, entonces, con probabilidad 1, la partícula alcanza el valor $z = k$. Una vez que lo alcanza, se inicia una caminata aleatoria del mismo tipo que la inicial pero partiendo de $z = k$, de manera que, por la proposición 7.39, la probabilidad de que esa caminata alcance el valor $z = k + 1$ es igual a 1. Por lo tanto $P[Z_{k+1} < \infty] = 1$.

Por el principio de inducción se puede concluir entonces que $P[Z_n < \infty] = 1$ para cualquier $n \in \mathbb{N}$. Por lo tanto, para cualquier $n \in \mathbb{N}$, con probabilidad 1, la partícula alcanza el valor $z = n$. Una vez que lo alcanza, se inicia una caminata aleatoria del mismo tipo que la inicial pero partiendo de $z = n$, de manera que, por la proposición 7.37, la probabilidad de que esa caminata visite el valor $z = n$ una infinidad de veces es igual a 1.

Finalmente, definamos los eventos:

B_n: La partícula pasa una infinidad de veces por la posición $z = n$.

B: La partícula pasa una infinidad de veces por la posición $z = n$ para cualquier $n \in \mathbb{N}$.

Se tiene entonces:

$P[B^c] = P[B_1^c \cup B_2^c \cup \cdots] \leq \sum_{j=1}^{\infty} P\left[B_j^c\right] = 0$

El caso $z = -n$ se demuestra de manera similar y el caso $z = 0$ está demostrado en la proposición 7.37. Combinando los tres casos se obtiene la conclusión deseada.

■

7.7.5. Distribución del tiempo de primer paso por un valor positivo. Sea Z el paso en el cual la partícula alcanza un valor positivo por primera vez, entonces, Z puede tomar únicamente valores impares y, si $m \in \mathbb{N}$, entonces:

$P[Z = 2m-1] = P[S_1 \leq 0, \ldots, S_{2m-2} \leq 0, S_{2m-1} > 0]$

Pero:

$P[S_1 \leq 0, \ldots, S_{2m-2} \leq 0, S_{2m-1} > 0] + P[S_1 \leq 0, \ldots, S_{2m-2} \leq 0, S_{2m-1} \leq 0]$

$= P[S_1 \leq 0, \ldots, S_{2m-2} \leq 0]$

Así que:

$P[S_1 \leq 0, \ldots, S_{2m-2} \leq 0, S_{2m-1} > 0]$

$= P[S_1 \leq 0, \ldots, S_{2m-2} \leq 0] - P[S_1 \leq 0, \ldots, S_{2m-2} \leq 0, S_{2m-1} \leq 0]$

$= P[S_1 \leq 0, \ldots, S_{2m-2} \leq 0] - P[S_1 \leq 0, \ldots, S_{2m-2} \leq 0, S_{2m} \leq 0]$

$= \frac{1}{2^{2m-2}}\binom{2m-2}{m-1} - \frac{1}{2^{2m}}\binom{2m}{m}$

Por lo tanto:

$P[Z = 2m-1] = \frac{1}{2^{2m-2}}\binom{2m-2}{m-1} - \frac{1}{2^{2m}}\binom{2m}{m} = \frac{1}{2^{2m}}\binom{2m}{m}\left[\frac{2m}{2m-1} - 1\right] = \frac{1}{2m-1}\frac{1}{2^{2m}}\binom{2m}{m}$

EJERCICIOS

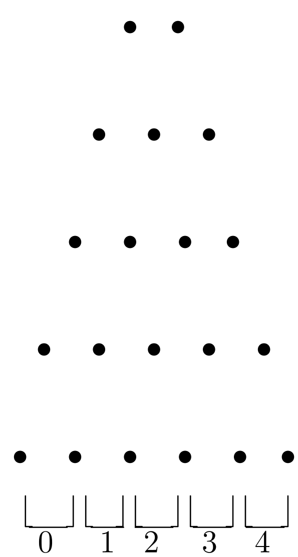
EJERCICIO 7.1. *Cada día el precio de una determinada acción sube o baja una unidad con probabilidades de* $\frac{3}{4}$ *y* $\frac{1}{4}$*, respectivamente. ¿Cuál es la probabilidad de que en* 6 *días el precio de la acción regresará a su valor original?*

EJERCICIO 7.2. *Asumiendo que la probabilidad de que un hijo procreado por una pareja sea niño es igual a* $\frac{1}{2}$*, ¿qué evento es más probable, que una pareja con* 4 *hijos tenga* 2 *hombres y* 2 *mujeres, o bien* 3 *hijos de un sexo y* 1 *del otro?*

EJERCICIO 7.3. *¿Qué evento es más probable, obtener por lo menos un* 6 *al lanzar* 6 *veces un dado, por lo menos* 2 *seises al lanzar* 12 *eces un dado, o bien por lo menos* 3 *seises al lanzar* 18 *veces un dado?*

EJERCICIO 7.4. *Se quiere transmitir un mensaje que consiste de uno de los dos dígitos,* 0 *ó* 1*. Debido a la estática, un dígito transmitido es recibido incorrectamente con probabilidad igual a* 0.2*. Para reducir la probabilidad de error, en lugar de transmitir el dígito* 0 *se trasmite la secuencia* 00000 *y en lugar del* 1 *se transmite* 11111*. Quien recibe el mensaje decide que el dígito que se transmitió fue aquel que recibió por lo menos* 3 *veces. ¿Cuál es la probabilidad de que el mensaje sea recibido correctamente?*

EJERCICIO 7.5. *Considere el dispositivo que se muestra en la figura de abajo, en el cual cada punto representa un clavo colocado verticalmente. Cuando se hace pasar una pelotita entre los dos clavos de la parte superior, ésta golpea el clavo del nivel siguiente y con probabilidad* $\frac{1}{2}$ *continúa su camino por el hueco de la derecha o el de la izquierda. Este proceso continúa hasta que la pelotita cae en alguna de las* 5 *cajas de la parte inferior. Si se realiza este proceso con* $1,000$ *pelotitas de manera independiente, ¿aproximadamente cuántas de ellas caerán en la caja número* 2*?*



Ejercicio 7.6. *Sea X una variable aleatoria con distribución binomial de parámetros n y p y sea k un número entero entre 0 y n. Encuentre el valor de p que hace que $P(X = k)$ sea un máximo.*

Ejercicio 7.7. *Sea X una variable aleatoria con distribución binomial de parámetros n y p. Demuestre que la probabilidad de que X tenga como valor un número impar es igual a $\frac{1}{2}[1 - (q - p)^n]$.*

Ejercicio 7.8. *Sea X una variable aleatoria con distribución binomial de parámetros n y p. Si se sabe que el valor de X es k, ¿cuál es la probabilidad de que ocurra cada uno de los $\binom{n}{k}$ arreglos de k éxitos y $n - k$ fracasos?*

Ejercicio 7.9. *Dos jugadores de ajedrez, A y B, juegan una serie de juegos con la condición de que A ganará la serie si gana 12 partidas antes de que B gane 6, de otra manera, B gana la serie (los empates no cuentan). Si la probabilidad de que A le gane una partida a B es igual a $\frac{2}{3}$ y la serie se interrumpe cuando A ha ganado 7 juegos y B 4, ¿quién merece ser nombrado el ganador?*

Ejercicio 7.10. *Un fumador tiene dos cajas con N cerillos cada una. Cada vez que necesita un cerillo, selecciona al azar una de las cajas y lo toma de ahí. Sea X el número de cerillos que le quedan en el momento en que encuentre que la caja que selecciona ya no tiene cerillos. Encuentre la distribución de X.*

Ejercicio 7.11. *Supongamos que, para $n \geq 1$, una pareja tiene n hijos con probabilidad igual a αp^n, en donde $0 < \alpha \leq \frac{1-p}{p}$ y $0 < p < 1$. Supongamos además que cada uno de los hijos es hombre o mujer con probabilidad igual a $\frac{1}{2}$. Para $k \in \{1, 2, \ldots\}$, ¿cuál es la probabilidad de que una pareja tenga por lo menos k hijos, de los cuales k exactamente sean mujeres?*

Ejercicio 7.12. *Sea X una variable aleatoria con distribución geométrica de parámetro p. Encuentre $P(X \geq x)$ para todos los enteros no negativos x.*

Ejercicio 7.13. *Sea X una variable aleatoria discreta que toma únicamente valores enteros no negativos y tal que, para cualquier entero no negativo n, se tiene:*

$$P[X = n \mid X > n - 1] = P[X = 0]$$

Demuestre que X tiene distribución geométrica.

Ejercicio 7.14. *Se quiere seleccionar a una persona de entre n. Para ello, cada una lanza una moneda y se conviene en que si alguna obtiene un resultado distinto al de las otras, entonces esa persona es la seleccionada; de otra manera se vuelven a lanzar las monedas bajo las mismas condiciones. Encuentre la probabilidad de que la persona sea seleccionada en el intento número k.*

Ejercicio 7.15. *Dos basketbolistas, A y B, juegan a lanzar por turnos una pelota sobre el aro, deteniéndose en el momento en que uno de ellos logre acertar. En cada uno de sus turnos, los jugadores A y B tienen probabilidades de acertar iguales a p_1 y p_2, respectivamente. Sean X y Y los números de tiros que realizan A y B, respectivamente, sin contar el último, hasta que se termina el juego. Asumiendo que todos los lanzamientos son independientes y que el primero en tirar es A, argumente, sin realizar cálculos, para mostrar que la distribución de Y es geométrica. Encuentre la función de densidad de X y el parámetro p de la distribución de Y.*

Ejercicio 7.16. *Sea X una variable aleatoria con distribución Poisson de parámetro λ. Demuestre que la probabilidad de que X tenga como valor un número impar es igual a $\frac{1}{2}\left[e^{\lambda} - e^{-\lambda}\right]$.*

Ejercicio 7.17. *Supongamos que el número de accidentes que tiene una persona en un año tiene distribución Poisson con parámetro λ, de tal manera que, en una cierta población, $\lambda = 2$ para el 60 % de personas y $\lambda = 3$ para el 40 %. Si X es el número de accidentes en un año de una persona seleccionada al azar, encuentre la distribución de X.*

Ejercicio 7.18. *Utilice la aproximación de Poisson para calcular la probabilidad de que en una caja de 100 cerillos se encuentren a lo más 2 cerillos defectuosos si 3 % de los cerillos que se producen son defectuosos.*

Ejercicio 7.19. *Los astrónomos estiman que en la Vía Láctea hay alrededor de 100 billones de estrellas las cuales tienen planetas a su alrededor. Sea p la probabilidad de que en un sistema solar determinado haya vida inteligente. ¿Para qué valores de p la probabilidad de que haya vida inteligente en algún sistema solar de la Vía Láctea es mayor que $\frac{1}{2}$?*

Ejercicio 7.20. *Supongamos que el número de huevos, X, que pone un insecto tiene distribución Poisson con parámetro λ y que la probabilidad de que un huevo se desarrolle es igual a p. Sea Y el número de huevos que se desarrollan. Demuestre que Y tiene distribución Poisson con parámetro λp.*

Ejercicio 7.21. *Sea X una variable aleatoria con distribución Poisson de parámetro λ. Para un número entero positivo k fijo, encuentre el valor de λ que hace que $P(X = k)$ sea un máximo.*

Ejercicio 7.22. *En una fábrica se empacan cajas de cereal con pasas mezclando primero, en un gran recipiente, el cereal con las pasas y de tal manera que cada caja contenga, en promedio, 12 pasas. Estime la probabilidad de que una caja de cereal, seleccionada al azar, contenga menos de 7 pasas.*

Ejercicio 7.23. *Un libro de 500 páginas contiene 500 errores de imprenta. Estime la probabilidad de que una página, seleccionada al azar, contenga 3 o más errores.*

EJERCICIO 7.24. *Un experimento aleatorio consiste en seleccionar, al azar y sin reemplazo, dos bolas de una caja que contiene* N *bolas marcadas con los números* $1, \ldots N$. *Sea* X *el mayor de los dos números de las bolas seleccionadas. Encuentre la función de densidad de* X.

EJERCICIO 7.25. *Un experimento aleatorio consiste en seleccionar, al azar y sin reemplazo,* 3 *números del conjunto* $\{1, \ldots, N\}$. *Sea* X *el menor de los números seleccionados,* Z *el mayor de ellos y* Y *el número intermedio. Encuentre las funciones de densidad de* X, Y *y* Z.

EJERCICIO 7.26. *Un estudiante va a presentar un examen, el cual consistirá de* 5 *preguntas seleccionadas al azar de entre* 10 *que el profesor les dejó estudiar. Si el estudiante prepara únicamente las respuestas de* 8 *de las* 10 *preguntas, ¿cuál es la probabilidad de que responda bien por lo menos* 4 *preguntas del examen?*

EJERCICIO 7.27. *Utilice la identidad* $(1+t)^{r+s} = (1+t)^r(1+t)^s$ *para demostrar que si la variable aleatoria* X *tiene distribución hipergeométrica, entonces:*

$$\sum_k P[X = k] = 1$$

EJERCICIO 7.28. *Supongamos que 5 personas, incluidas A y B, se sientan al azar en 5 sillas colocadas en hilera y sea* X *el número de personas colocadas entre A y B. Encuentre la función de densidad de* X.

EJERCICIO 7.29. *Sea* X *una variable aleatoria con función de densidad dada por*

$$f(x) = \begin{cases} \frac{c}{3^x} & si\ x \in \{2, 3, \ldots\} \\ 0 & en\ otro\ caso \end{cases}$$

en donde c es una constante. Encuentre el valor de c y la probabilidad de que X *tome como valor un número par.*

EJERCICIO 7.30. *Demuestre que la función dada por:*

$$f_X(x) = \begin{cases} -\frac{1}{\ln(1-q)}\frac{q^x}{x} & si\ x \in \{1, 2, \ldots\} \\ 0 & en\ otro\ caso \end{cases}$$

es una función de densidad, en donde $q \in (0, 1]$.

EJERCICIO 7.31. *Sea* X *una variable aleatoria uniformemente distribuida en el conjunto* $\{0, 1, ..., 99\}$. *Calcule :*

a) $P(X \geq 25)$

b) $P(2.6 \leq X \leq 12.2)$

c) $P(8 < X \leq 10 \ \text{ó} \ 30 < X \leq 32)$

d) $P(25 \leq X \leq 30)$

EJERCICIO 7.32. *Un experimento aleatorio consiste en seleccionar al azar un número entero entre* 0 *y* 99 *(inclusive). Sea* X *la suma de los dígitos que se obtienen. Encuentre la función de densidad de* X.

EJERCICIO 7.33. *Sea* X *y* Y *dos variables aleatorias independientes,* X *con distribución uniforme en el conjunto* $\{1,\ldots,N\}$*,* Y *con distribución uniforme en el conjunto* $\{1,\ldots,M\}$*, en donde* $N \leq M$*. Encuentre* $P[X < Y]$ *y* $P[X = Y]$.

EJERCICIO 7.34. *Sea* X *una variable aleatoria distribuida uniformemente en el conjunto* $\{1,\ldots,N\}$*, en donde* N *es un entero positivo y sea* $M < N$ *otro entero positivo. Encuentre la función de densidad de* $Y = \text{máx}(X, M)$.

EJERCICIO 7.35. *Sea* X *una variable aleatoria distribuida uniformemente en el conjunto* $\{1,\ldots,N\}$*, en donde* N *es un entero positivo y sea* $1 < M \leq N$ *otro entero positivo. Encuentre la función de densidad de* $Y = \text{mín}(X, M)$.

EJERCICIO 7.36. *Sean* X *y* Y *dos variables aleatorias independientes, ambas con distribución uniforme en el conjunto* $\{1,\ldots,N\}$*. Encuentre la función de densidad de* $Z = \text{máx}(X, Y)$.

EJERCICIO 7.37. *Sean* X *y* Y *dos variables aleatorias independientes,* X *con distribución geométrica de parámetro* p*,* Y *con distribución Poisson de parámetro* λ*. a) Encuentre* $P[X < Y]$ *y* $P[X > Y]$*. b) Determine los valores de* p *y* λ *para los cuales* $P[X > Y] > P[X < Y]$.

EJERCICIO 7.38. *Sean* X *y* Y *dos variables aleatorias independientes, ambas con distribución geométrica de parámetro* p*. Encuentre a)* $P[X < Y]$ *y* $P[X = Y]$ *y b) la función de densidad de* $Z = \text{mín}(X, Y)$.

CAPÍTULO 8

VARIABLES ALEATORIAS ABSOLUTAMENTE CONTINUAS

La doctrina de los azares puede ser una ayuda para curar una especie de superstición, la cual ha estado arraigada por mucho tiempo en el mundo, a saber, que hay en los juegos algo como la suerte, buena o mala.

Abraham de Moivre

8.1. Distribución uniforme continua

DEFINICIÓN 8.1 (**Distribución uniforme en** $(\mathbf{a}, \mathbf{b})$). *Se dice que una variable aleatoria X tiene distribución uniforme en el intervalo (a, b) si la función:*

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \textit{si } x \in (a, b) \\ 0 & \textit{en otro caso} \end{cases}$$

es una función de densidad de X.

Una distribución uniforme en el intervalo $(0, 1)$ no es otra cosa que la medida de Lebesgue en ese intervalo (ver sección 5.7). En efecto, sea X una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$ y J un intervalo (de cualquier tipo), de extremos c y d, contenido en el intervalo $(0, 1)$, entonces:

$$P[X \in J] = \int_c^d dx = d - c = l(J)$$

en donde $l(J)$ denota la longitud del intervalo J.

Pero, de acuerdo con la proposición 5.52, la medida de Lebesgue es la única medida de probabilidad, definida sobre los conjuntos Lebesgue medibles contenidos en el intervalo $(0,1)$, que asigna a cada intervalo su longitud. Por lo tanto, si m denota a la medida de Lebesgue, se tiene $P[X \in A] = m(A)$ para cualquier conjunto Lebesgue medible contenido en el intervalo $(0,1)$.

La distribución uniforme continua surge como una extensión al caso continuo del concepto de equiprobabilidad. Se encuentra esta distribución, por ejemplo, en un artículo de Thomas Bayes publicado en el año 1763 ("An essay towards solving a problem in the doctrine of chances", Philos. Trans. Roy. Soc. London Ser. A 53, 1763, reproducido en Biometrika 45, 1958) en el cual, dado un rectángulo ABCD, con base el segmento AB y altura el segmento AC, Bayes consideró el experimento aleatorio que consiste en lanzar una bola W sobre el rectángulo, de tal manera que haya la misma probabilidad de que la bola caiga en cualquiera de dos partes iguales del rectángulo.

C D

A s B

A través del punto en donde cae la bola W, se traza una recta paralela a AC, la cual corta al segmento AB en el punto s. Encontró entonces que la probabilidad de que el punto s caiga entre cualquier par de puntos del segmento AB es igual a la razón de la distancia entre los dos puntos a la longitud del segmento AB. Es decir, mostró que la distribución del punto s en el segmento AB es uniforme.

De manera general, la distribución uniforme se obtiene al considerar el experimento aleatorio consistente en la elección al azar de un número real en el intervalo (a,b) y definiendo X como el número que se obtiene. La. elección al azar se interpreta en el sentido de que dos subconjuntos del intervalo (a,b) que se puedan sobreponer geométricamente uno sobre el otro deben de tener asignada la misma probabilidad.

Ya estudiamos este tipo de experimentos en la sección 5.2, y ahí encontramos que si J es un intervalo de cualquier tipo, de extremos c y d, contenido en (a,b), entonces:

$P[X \in J] = \frac{d-c}{b-a}$

Por lo tanto:

$$F_X(x) = P[X \leq x] = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } x \geq b \end{cases}$$

Así que, una función de densidad de X está dada por:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in (a,b) \\ 0 & \text{en otro caso} \end{cases}$$

EJEMPLO 8.2. *Consideremos el experimento aleatorio consistente en la elección al azar de un número real en el intervalo $(0,1)$ y sea X el número que se obtiene. ¿Cuál es la probabilidad de que a) X sea un número racional? y b) X pertenezca al conjunto de Cantor?*

Solución

a) Como $P[X = x] = 0$ para cualquier $x \in \mathbb{R}$ y el conjunto de los números racionales es numerable, la propiedad de la aditividad numerable nos permite concluir que la probabilidad de que X sea un número racional es igual a 0.

b) Sea $\mathcal{C}$ el conjunto de Cantor, entonces:

$$[0,1] - \mathcal{C} = \left(\tfrac{1}{3}, \tfrac{2}{3}\right) \cup \left(\tfrac{1}{3^2}, \tfrac{2}{3^2}\right) \cup \left(\tfrac{7}{3^2}, \tfrac{8}{3^2}\right) \cup \left(\tfrac{1}{3^3}, \tfrac{2}{3^3}\right) \cup \left(\tfrac{7}{3^3}, \tfrac{8}{3^3}\right) \cup \left(\tfrac{19}{3^3}, \tfrac{20}{3^3}\right) \cup \left(\tfrac{25}{3^3}, \tfrac{26}{3^3}\right) \cup \cdots$$

Así que:

$$P[X \in \mathcal{C}] = 1 - P[X \in [0,1] - \mathcal{C}]$$

$$= 1 - P\left[X \in \left(\tfrac{1}{3}, \tfrac{2}{3}\right)\right] - P\left[X \in \left(\tfrac{1}{3^2}, \tfrac{2}{3^2}\right)\right] - P\left[X \in \left(\tfrac{7}{3^2}, \tfrac{8}{3^2}\right)\right] - \cdots$$

$$= 1 - \left(\tfrac{1}{3} + \tfrac{2}{3^2} + \tfrac{2^2}{3^3} + \tfrac{2^3}{3^4} + \cdots\right) = 1 - \frac{\frac{1}{3}}{\frac{1}{3} - \frac{2}{3}} = 0$$

▲

La distribución uniforme puede también obtenerse dentro del contexto de los ensayos de Bernoulli. Supongamos que se realiza una sucesión infinita de ensayos de Bernoulli independientes, en cada uno de los cuales la probabilidad de éxito es igual a $\frac{1}{2}$. Los posibles resultados de este experimento aleatorio son entonces sucesiones infinitas (c_k) de éxitos y fracasos, los cuales pueden ser representados por 1's y 0's, respectivamente. Si $\omega = (c_k)$, definamos $X(\omega) = \sum_{k=1}^{\infty} \frac{c_k}{2^k}$. En otras palabras, cada sucesión (c_k) se está viendo como el desarrollo en base 2 de un número real en el intervalo $[0,1]$ y $X(\omega)$ es precisamente ese número real.

Para obtener la distribución de X, observemos primero que $P[X=x]=0$ para cualquier número real x. En efecto, llamemos X_k al resultado del k-ésimo ensayo de Bernoulli. Entonces, dado cualquier posible resultado $\omega=(c_k)$, se tiene, para cualquier $n\in\mathbb{N}$:

$P[\{\omega\}]=P[X_1=c_1,X_2=c_2,\ldots,X_n=c_n,\ldots]$

$\leq P[X_1=c_1,X_2=c_2,\ldots,X_n=c_n]=P[X_1=c_1]\,P[X_2=c_2]\cdots P[X_n=c_n]=\frac{1}{2^n}$

Por lo tanto, $P[\{\omega\}]=0$.

Ahora bien, dado un número real $x\in[0,1]$, el evento $[X=x]$ ocurre si y sólo si se obtiene como resultado una sucesión $\omega=(c_k)$, la cual representa al número x escrito en base 2. Pero cada punto x tiene a lo más dos desarrollos en base 2, es decir, el evento $[X=x]$ está formado por a los más dos posibles resultados del espacio muestral, cada uno de los cuales tiene probabilidad 0. Por lo tanto $P[X=x]=0$.

Para cada $n\in\mathbb{N}$, consideremos ahora un intervalo de la forma $(\frac{k}{2^n},\frac{k+1}{2^n})$, con $0\leq k<2^n$. Asociada a tal intervalo existe una única colección de n 0's y 1's, $c_1,c_2,\ldots,c_n$, tal que un punto x pertenece al intervalo $(\frac{k}{2^n},\frac{k+1}{2^n})$ si y sólo si tiene un desarrollo en base 2 de la forma $x=0.c_1c_2\ldots c_n\ldots$. Por lo tanto:

$P\left[X\in(\frac{k}{2^n},\frac{k+1}{2^n})\right]=P[X_1=c_1,X_2=c_2,\ldots,X_n=c_n]$

$=P[X_1=c_1]\,P[X_2=c_2]\cdots P[X_n=c_n]=\frac{1}{2^n}$

Es decir, la probabilidad del evento $\left[X\in(\frac{k}{2^n},\frac{k+1}{2^n})\right]$ es igual a la longitud del intervalo $(\frac{k}{2^n},\frac{k+1}{2^n})$.

Finalmente, dado un intervalo $(a,b)\subset(0,1)$ y $\varepsilon>0$, sea $n\in\mathbb{N}$ tal que $\frac{1}{2^n}<\text{mín}\left\{\varepsilon,\frac{b-a}{4},a,1-b\right\}$, r el más pequeño número entero no negativo k tal que $\frac{k}{2^n}\in(a,b)$ y M el más grande número entero no negativo k tal que $\frac{r+k}{2^n}\in(a,b)$. Consideremos los intervalos $I_1=(\frac{r}{2^n},\frac{r+1}{2^n})$, $I_2=(\frac{r+1}{2^n},\frac{r+2}{2^n})$, ..., $I_M=(\frac{r+M-1}{2^n},\frac{r+M}{2^n})$. Entonces (a,b) y $(a,\frac{r}{2^n})\cup I_1\cup I_2\cup\cdots\cup I_M\cup(\frac{r+M}{2^n},b)$ difieren únicamente por un número finito de puntos, así que:

$P[X\in(a,b)]=\sum_{k=1}^{M}\ell(I_k)+P\left[X\in(a,\frac{r}{2^n})\right]+P\left[X\in(\frac{r+M}{2^n},b)\right]$

$\ell[(a,b)]=\sum_{k=1}^{M}\ell(I_k)+\ell\left[(a,\frac{r}{2^n})\right]+\ell\left[(\frac{r+M}{2^n},b)\right]$

Además:

$0\leq P\left[X\in(a,\frac{r}{2^n})\right]+P\left[X\in(\frac{r+M}{2^n},b)\right]$

$\leq P\left[X \in (\frac{r-1}{2^n}, \frac{r}{2^n})\right] + P\left[X \in (\frac{r+M}{2^n}, \frac{r+M+1}{2^n})\right] = \frac{2}{2^n} < 2\varepsilon$

$0 \leq \ell\left[(a, \frac{r}{2^n})\right] + \ell\left[(\frac{r+M}{2^n}, b)\right] < \frac{2}{2^n} < 2\varepsilon$

Así que:

$0 \leq P\left[X \in (a,b)\right] - \sum_{k=1}^{M} \ell(I_k) < 2\varepsilon$

$0 \leq \ell\left[(a,b)\right] - \sum_{k=1}^{M} \ell(I_k) < 2\varepsilon$

Por lo tanto:

$\left|P\left[X \in (a,b)\right] - \ell\left[(a,b)\right]\right| < 2\varepsilon$

Como esto es válido para cualquier $\varepsilon > 0$, se puede concluir que:

$P\left[X \in (a,b)\right] = \ell\left[(a,b)\right]$

8.2. Distribución normal

DEFINICIÓN 8.3 (**Distribución normal**). *Sean $\mu, \sigma \in \mathbb{R}$, con $\sigma > 0$. Se dice que una variable aleatoria X tiene distribución normal con parámetros μ y σ^2 si la función:*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

es una función de densidad de X.

Verifiquemos que esta función, así definida, es una función de densidad:

$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-y^2} dy = 1$

La integral indefinida $\int e^{-y^2} dy$ no se puede expresar en forma cerrada, de ahí que para evaluar una integral definida $\int_a^b e^{-y^2} dy$ se tenga que recurrir a métodos numéricos. Por esta razón es común encontrar tablas que contienen evaluaciones de ese tipo de integrales para diferentes intervalos. Con el objeto de no tener que recurrir a una tabla diferente para cada valor de μ y σ, para evaluar una integral de la forma $\int_a^b \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx$, es conveniente reducir ésta, mediante un cambio de variable, a una de la forma $\int_a^b e^{-\frac{1}{2}y^2} dy$. De manera equivalente, una variable aleatoria con distribución normal de parámetros μ y σ^2 se puede transformar, de manera muy simple, en una con distribución normal con parámetros 0 y 1. Esto se formula en la siguiente proposición:

Proposición 8.4. *Sea X una variable aleatoria con distribución normal de parámetros μ y σ^2, entonces $Y = \frac{X-\mu}{\sigma}$ tiene distribución normal con parámetros* 0 *y* 1*.*

Demostración

$F_Y(y) = P[Y \leq y] = P\left[\frac{X-\mu}{\sigma} \leq y\right] = P[X \leq \mu + \sigma y] = F_X(\mu + \sigma y)$

Por lo tanto, una función de densidad de Y está dada por:

$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(\mu + \sigma y) = \sigma f_X(\mu + \sigma y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\}$

■

Definición 8.5 (**Distribución normal estándar**). *Se dice que una variable aleatoria X tiene distribución normal estándar si su distribución es normal con parámetros $\mu = 0$ y $\sigma^2 = 1$.*

Dada una variable aleatoria X, con distribución normal, la proposición 8.4 permite calcular cualquier probabilidad de la forma $P[a < X < b]$, mediante una tabla correspondiente a la distribución normal estándar.

Otra simplificación se obtiene al considerar que la función $e^{-\frac{1}{2}y^2}$ es simétrica con respecto al eje vertical, lo cual implica, en particular, que:

$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{1}{2}y^2} dy = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{1}{2}y^2} dy = 0.5$

$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-a} e^{-\frac{1}{2}y^2} dy = \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-\frac{1}{2}y^2} dy$ para cualquier $a > 0$

$\frac{1}{\sqrt{2\pi}} \int_{-a}^{0} e^{-\frac{1}{2}y^2} dy = \frac{1}{\sqrt{2\pi}} \int_{0}^{a} e^{-\frac{1}{2}y^2} dy$ para cualquier $a > 0$

En la página 403 se presenta una tabla de la distribución normal estándar. Nótese que ésta contiene las evaluaciones de integrales de la forma $\frac{1}{\sqrt{2\pi}} \int_{0}^{a} e^{-\frac{1}{2}y^2} dy$ para $a \in \{0.00,$ 0.01, 0.02, $\ldots$, $3.90\}$.

Para $z \geq 0$, definamos:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{0}^{z} e^{-\frac{1}{2}y^2} dy$$

Para valores de $z \in (0, 3.9)$, que no pertenezcan al conjunto $A = \{0.00, 0.01, \ldots, 3.90\}$, se puede obtener una aproximación de $\Phi(z)$ a partir de los valores de Φ en los dos puntos, $z_1, z_2 \in A$ entre los cuales se encuentra z. Para esto, se considera la recta en $\mathbb{R}^2$ que une los puntos $(z_1, \Phi(z_1))$ con $(z_2, \Phi(z_2))$; si (z, u) se encuentra sobre esa recta, entonces u es una aproximación de $\Phi(z)$. Por ejemplo:

$\Phi(1.367) \approx \Phi(1.36) + \frac{\Phi(1.37)-\Phi(1.36)}{1.37-1.36}(1.367 - 1.36)$

$= 0.4131 + \frac{0.4147-0.4131}{1.37-1.36}(1.367 - 1.36) = 0.41422$

Ejemplo 8.6. *Sea X una variable aleatoria con distribución normal de parámetros $\mu = 2.4$ y $\sigma^2 = 1.6$. Encuentre $P[2 < X < 3]$, $P[-1 < X < 0.5]$, $P[X > -1.5]$ y $P[X > 3.1]$.*

Solución

$P[2 < X < 3] = P\left[\frac{2-2.4}{\sqrt{1.6}} < \frac{X-\mu}{\sigma} < \frac{3-2.4}{\sqrt{1.6}}\right] = P\left[-0.316 < \frac{X-\mu}{\sigma} < 0.474\right]$

$= \Phi(0.316) + \Phi(0.474) \approx 0.12398 + 0.18224 = 0.30622$

$P[-1 < X < 0.5] = P\left[\frac{-1-2.4}{\sqrt{1.6}} < \frac{X-\mu}{\sigma} < \frac{0.5-2.4}{\sqrt{1.6}}\right] = P\left[-2.688 < \frac{X-\mu}{\sigma} < -1.502\right]$

$= \Phi(2.688) - \Phi(1.502) \approx 0.49638 - 0.43346 = 0.06292$

$P[X > -1.5] = P\left[\frac{X-\mu}{\sigma} > \frac{-1.5-2.4}{\sqrt{1.6}}\right] = P\left[\frac{X-\mu}{\sigma} > -3.083\right] = 0.5 + \Phi(3.083)$

$= 0.5 + 0.499 = 0.999$

$P[X > 3.1] = P\left[\frac{X-\mu}{\sigma} > \frac{3.1-2.4}{\sqrt{1.6}}\right] = P\left[\frac{X-\mu}{\sigma} > 0.553\right] = 0.5 - \Phi(0.553)$

$\approx 0.5 - 0.20985 = 0.29015$

▲

La distribución normal es una de las distribuciones más importantes, tanto en la Estadística como en la Teoría de la Probabilidad, pues aproxima a muchas otras distribuciones. De hecho, surgió como una distribución límite, siendo Abraham de Moivre el primero en obtenerla como una aproximación de la distribución binomial. Esto lo hizo en su trabajo titulado "Approximatio ad summam terminorum binomii (a+b)n in seriem expansi", publicado en 1733, en el cual investigó las desviaciones que se presentan entre la frecuencia relativa de ocurrencia de un evento A en n repeticiones del experimento correspondiente y la probabilidad de ocurrencia de A. Más tarde Laplace complementó ese trabajo, por lo cual actualmente se le conoce como el teorema de de Moivre-Laplace.

El resultado de de Moivre-Laplace constituye el segundo de los llamados **teoremas límite** de la Teoría de la Probabilidad. Este resultado y el de Bernoulli (ver proposición 7.4) marcaron la pauta para el desarrollo posterior de la Teoría de la Probabilidad, concluyéndose su estudio apenas a principios del siglo XX con la formulación general de los teoremas límite.

Dentro de la teoría moderna, el teorema de de Moivre-Laplace es únicamente un caso particular del teorema del límite central. Sin embargo, resulta ilustrativa una demostración directa, la cual puede realizarse con la herramienta que se tiene desarrollada hasta este momento.

8.3. Teorema de de Moivre-Laplace

TEOREMA 8.7 (**Teorema local de de Moivre-Laplace**). *Para cada $n \in \mathbb{N}$, sea X_n una variable aleatoria con distribución binomial de parámetros n y $p \in (0,1)$. Dados a y b dos números reales tales que $a < b$, definamos, para cada entero no negativo k tal que $np + a\sqrt{npq} \leq k \leq np + b\sqrt{npq}$, $x_{k,n} = \frac{k-np}{\sqrt{npq}}$, entonces:*

$$\lim_{n \rightsquigarrow \infty} \frac{P[X_n=k]}{\frac{1}{\sqrt{2\pi}\sqrt{npq}} \exp\left\{-\frac{1}{2}x_{k,n}^2\right\}} = 1 \textit{ uniformemente en } k.$$

Demostración

Por la fórmula de Stirling, se tiene, para cualquier $m \in \mathbb{N}$:

$m! = \sqrt{2\pi} m^m e^{-m} \sqrt{m} e^{\theta m}$, en donde $\frac{1}{12m+1} < \theta_m < \frac{1}{12m}$.

Por lo tanto, para $k \in \{0, 1, \ldots, n\}$, se tiene:

$P[X_n = k] = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$

$= \frac{n^n \sqrt{n} e^{\theta_n}}{\sqrt{2\pi} k^k \sqrt{k} e^{\theta_k} (n-k)^{n-k} \sqrt{n-k} e^{\theta_{n-k}}} p^k (1-p)^{n-k}$

$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} \sqrt{\frac{np}{k}} \sqrt{\frac{nq}{n-k}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} e^{\theta_n - \theta_k - \theta_{n-k}}$

Ahora bien, si k satisface $np + a\sqrt{npq} \leq k \leq np + b\sqrt{npq}$, entonces $nq - b\sqrt{npq} \leq n - k \leq nq - a\sqrt{npq}$, así que:

$\frac{np}{np+b\sqrt{npq}} \leq \frac{np}{k} \leq \frac{np}{np+a\sqrt{npq}}$

$\frac{nq}{nq-a\sqrt{npq}} \leq \frac{np}{k} \leq \frac{np}{nq-b\sqrt{npq}}$

$\theta_n - \theta_k - \theta_{n-k} \leq \theta_n + \theta_k + \theta_{n-k} < \frac{1}{12}\left(\frac{1}{n} + \frac{1}{k} + \frac{1}{n-k}\right) \leq \frac{1}{12}\left(\frac{1}{n} + \frac{1}{np+a\sqrt{npq}} + \frac{1}{nq-b\sqrt{npq}}\right)$

Por lo tanto:

$\lim_{n \rightsquigarrow \infty} \sqrt{\frac{np}{k}} \sqrt{\frac{nq}{n-k}} e^{\theta_n - \theta_k - \theta_{n-k}} = 1$ uniformemente en k.

Sea ahora $H_n(k) = \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}$

Obsérvese entonces que:

$\frac{k}{np} = \frac{np+x_{k,n}\sqrt{npq}}{np} = 1 + x_{k,n}\sqrt{\frac{q}{np}}$

$\frac{n-k}{nq} = \frac{nq-x_{k,n}\sqrt{npq}}{nq} = 1 - x_{k,n}\sqrt{\frac{p}{nq}}$

Así que:

$\ln H_n(k) = -(np + x_{k,n}\sqrt{npq}) \ln\left(1 + x_{k,n}\sqrt{\frac{q}{np}}\right) - (nq - x_{k,n}\sqrt{npq}) \ln\left(1 - x_{k,n}\sqrt{\frac{p}{nq}}\right)$

Pero, existen C y N tales que:

$\ln\left(1 + x\sqrt{\frac{q}{np}}\right) = x\sqrt{\frac{q}{np}} - \frac{1}{2}\frac{q}{np}x^2 + A_n(x)$

$\ln\left(1 - x\sqrt{\frac{p}{nq}}\right) = -x\sqrt{\frac{p}{nq}} - \frac{1}{2}\frac{p}{nq}x^2 + B_n(x)$

en donde $n\sqrt{n}\,|A_n(x)| < C$, $n\sqrt{n}\,|B_n(x)| < C$ para $n \geq N$ y $x \in [a, b]$.

Por lo tanto:

$\ln H_n(k) = -(np + x_{k,n}\sqrt{npq}) \left[x_{k,n}\sqrt{\frac{q}{np}} - \frac{1}{2}\frac{q}{np}x_{k,n}^2 + A_n(x_{k,n})\right]$

$-(nq - x_{k,n}\sqrt{npq}) \left[-x_{k,n}\sqrt{\frac{p}{nq}} - \frac{1}{2}\frac{p}{nq}x_{k,n}^2 + B_n(x_{k,n})\right]$

$= -\frac{1}{2}x_{k,n}^2 - (np + x_{k,n}\sqrt{npq})A_n(x_{k,n}) - (nq - x_{k,n}\sqrt{npq})B_n(x_{k,n})$

Ahora bien:

$\left|(np + x_{k,n}\sqrt{npq})A_n(x_{k,n})\right| = \left|\left(\frac{p}{\sqrt{n}} + \frac{x_{k,n}}{n}\sqrt{pq}\right) n\sqrt{n}A_n(x_{k,n})\right| < C\left|\frac{p}{\sqrt{n}} + \frac{x_{k,n}}{n}\sqrt{pq}\right|$

Así que:

$\lim_{n\rightsquigarrow\infty} \left|(np + x_{k,n}\sqrt{npq})A_n(x_{k,n})\right| = 0$ uniformemente en k.

De la misma manera:

$\lim_{n\rightsquigarrow\infty} \left|(nq - x_{k,n}\sqrt{npq})B_n(x_{k,n})\right| = 0$ uniformemente en k.

Por lo tanto, si $G_n(k) = \exp\left[-(np + x_{k,n}\sqrt{npq})A_n(x_{k,n}) - (nq - x_{k,n}\sqrt{npq})B_n(x_{k,n})\right]$, entonces $\lim_{n\rightsquigarrow\infty} G_n(k) = 1$ uniformemente en k y $H_n(k) = G_n(x_{k,n})e^{-\frac{1}{2}x_{k,n}^2}$.

Así que:

$\lim_{n\rightsquigarrow\infty} \sqrt{2\pi}\sqrt{npq}P\left[X_n = k\right] e^{\frac{1}{2}x_{k,n}^2}$

$= \lim_{n\rightsquigarrow\infty} \frac{1}{\sqrt{2\pi}}\sqrt{\frac{np}{k}}\sqrt{\frac{nq}{n-k}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k} e^{\theta_n-\theta_k-\theta_{n-k}} e^{\frac{1}{2}x_{k,n}^2}$

$= \lim_{n\rightsquigarrow\infty} \sqrt{\frac{np}{k}}\sqrt{\frac{nq}{n-k}}e^{\theta_n-\theta_k-\theta_{n-k}}H_n(k)e^{\frac{1}{2}x_{k,n}^2} = \lim_{n\rightsquigarrow\infty} \sqrt{\frac{np}{k}}\sqrt{\frac{nq}{n-k}}e^{\theta_n-\theta_k-\theta_{n-k}}G_n(k)$

$= \lim_{n\rightsquigarrow\infty} \left(\sqrt{\frac{np}{k}}\sqrt{\frac{nq}{n-k}}e^{\theta_n-\theta_k-\theta_{n-k}}\right)\lim_{n\rightsquigarrow\infty} G_n(k) = 1$ uniformemente en k.

∎

Corolario 8.8. *Para cada $n \in \mathbb{N}$, sea X_n una variable aleatoria con distribución binomial de parámetros n y $p \in (0,1)$ y k un entero no negativo tal que $k \leq n$. Entonces:*

$\lim_{n\rightsquigarrow\infty} P\left[X_n = k\right] = 0$ *uniformemente en k*

Demostración

Recordemos que si X es una variable aleatoria con distribución binomial de parámetros n y p, entonces los términos $P\left[X = k\right]$ crecen con k hasta alcanzar su máximo valor cuando $np + p - 1 \leq k \leq np + p$, después de lo cual decrecen con k. Entonces, para probar el resultado, basta con demostrar que el término máximo tiende a cero cuando $n \rightsquigarrow \infty$. Para cada $n \in \mathbb{N}$, sea k_n un entero no negativo tal que:

$P\left[X_n = k_n\right] = \max\left\{P\left[X_n = k\right] : k \text{ es un entero no negativo tal que } k \leq n\right\}$

Entonces, tomando n suficientemente grande, se tiene $np - \sqrt{npq} \leq k_n \leq np + \sqrt{npq}$. Así que, por el teorema local de de Moivre-Laplace, si $x_n = \frac{k_n - np}{\sqrt{npq}}$, entonces:

$\lim_{n\rightsquigarrow\infty} \frac{\sqrt{2\pi}\sqrt{npq}P[X_n=k_n]}{\exp\left\{-\frac{1}{2}x_n^2\right\}} = 1$

Pero $p-1 \leq k_n - np \leq p$, así que $\lim_{n\rightsquigarrow\infty} x_n = 0$. Por lo tanto, $\lim_{n\rightsquigarrow\infty} \exp\left\{-\frac{1}{2}x_n^2\right\} = 1$ y entonces:

$\lim_{n\rightsquigarrow\infty} \sqrt{2\pi}\sqrt{npq}P\left[X_n = k_n\right] = 1$

lo cual implica que $\lim_{n\rightsquigarrow\infty} P\left[X_n = k_n\right] = 0$.

∎

El teorema local del límite puede utilizarse para encontrar aproximaciones de la función de densidad de una variable aleatoria con distribución binomial. Para esto, en la búsqueda de una estimación de una probabilidad $P[X = k]$, se define $x_{k,n} = \frac{k-np}{\sqrt{npq}}$. De esta forma, se tendrá $P[X = k] \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_{k,n}^2}$. Es decir:

$$P[X = k] \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{k-np}{\sqrt{npq}} \right)^2 \right\}$$

Obsérvese ahora que la función $f(x) = \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{x-np}{\sqrt{npq}} \right)^2 \right\}$ es una función de densidad de una variable aleatoria con distribución normal de parámetros $\mu = np$ y $\sigma^2 = npq$. De esta manera se tiene el siguiente resultado:

Corolario 8.9. *Sea X una variable aleatoria con distribución binomial de parámetros n y $p \in [0, 1]$ y $x \in \{0, \ldots, n\}$, entonces $P[X = x] \approx f(x)$, en donde f es una función de densidad de una variable aleatoria con distribución normal de parámetros $\mu = np$ y $\sigma^2 = npq$.*

De acuerdo con el teorema local del límite, la aproximación será mejor cuando k esté más cerca de np.

Ejemplo 8.10. *Sea X una variable aleatoria con distribución binomial de parámetros $n = 50$ y $p = \frac{1}{4}$. Entonces, de acuerdo con el último corolario, para $x \in \{0, \ldots, 50\}$, se tiene:*

$f_X(x) = P[X = x] = \binom{50}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{50-x}$

$f_X(x) \approx \frac{2}{5}\sqrt{\frac{1}{3\pi}} \exp\left\{ -\frac{4}{75} \left(x - \frac{25}{2}\right)^2 \right\}$

Con el objeto de comparar los valores exactos de una función de densidad binomial con sus valores aproximados por medio de una densidad normal, a continuación se muestran las gráficas que se obtienen con ambos valores. Con el objeto de apreciar mejor la aproximación, la gráfica de los valores exactos de la densidad binomial f_X se ha trazado para todos los valores reales de x comprendidos en el intervalo $[0, 50]$. Para esto se ha expresado f_X en la forma:

$f_X(x) = \frac{50!}{\Gamma(x+1)\Gamma(51-x)} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{50-x}$

La línea sólida corresponde a los valores exactos, mientras que la línea punteada corresponde a los valores aproximados.

De acuerdo con las gráficas, la aproximación es bastante buena y pareciera que es mejor en las colas de la distribución que en el centro de ella. Sin embargo, esto no es así pues si bien el error absoluto que se comete con la aproximación es más grande en el centro de la distribución, el error relativo es menor en esa misma región. Por ejemplo, se tiene:

$P[X=2] = \binom{50}{2}(.25)^2(.75)^{48} = 7.7083 \times 10^{-5}$

$P[X=10] = \binom{50}{10}(.25)^{10}(.75)^{40} = 0.09\,8518$

Mientras que, utilizando el teorema de de Moivre-Laplace, se tiene:

$P[X=2] \approx \frac{2}{5}\sqrt{\frac{1}{3\pi}}\exp\left\{-\frac{4}{75}\left(2-\frac{25}{2}\right)^2\right\} = 3.6414 \times 10^{-4}$

$P[X=10] \approx \frac{2}{5}\sqrt{\frac{1}{3\pi}}\exp\left\{-\frac{4}{75}\left(10-\frac{25}{2}\right)^2\right\} = 0.09336$

Para el caso de $P[X=2]$, los errores absolutos y relativos que se cometen con la aproximación están dados, respectivamente, por:

$e_a = 3.\,6414 \times 10^{-4} - 7.\,7083 \times 10^{-5} = 2.8706 \times 10^{-4}$

$e_r = \frac{e_a}{7.7083\times 10^{-5}} = \frac{2.8706\times 10^{-4}}{7.7083\times 10^{-5}} = 3.724$

Por otro lado, para el caso de $P[X=10]$, se tiene:

$e_a = 0.09336 - 0.098518 = -5.158 \times 10^{-3}$

$e_r = \frac{e_a}{0.098518} = \frac{-5.158\times 10^{-3}}{.098518} = -0.052356$

Entonces, efectivamente, en el primer caso, el error absoluto es menor, pero el error relativo es mucho mayor.

▲

Cuando n no es grande, la aproximación no es muy buena. Por ejemplo, a continuación se presentan las gráficas de los valores exactos y aproximados para el caso de una variable aleatoria con distribución binomial de parámetros $n = 12$ y $p = \frac{1}{4}$.

Teorema 8.11 (**Teorema integral de de Moivre-Laplace**). *Sean a y b dos números reales tales que $a < b$ y, para cada $n \in \mathbb{N}$, sea X_n una variable aleatoria con distribución binomial de parámetros n y $p \in (0, 1)$. Entonces:*

$$\text{lím}_{n \rightsquigarrow \infty} P\left[a < \frac{X_n - np}{\sqrt{npq}} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx$$

Demostración

Para cada entero no negativo k tal que $a < \frac{k-np}{\sqrt{npq}} < b$, definamos $x_{k,n} = \frac{k-np}{\sqrt{npq}}$ y $\alpha(n,k) = \frac{\sqrt{2\pi}\sqrt{npq}P[X_n=k]}{\exp\left\{-\frac{1}{2}x_{k,n}^2\right\}} - 1$, entonces:

$$P\left[X_n = k\right] = \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x_{k,n}^2\right\}\left[1 + \alpha(n,k)\right]$$

Así que:

$$P\left[a < \frac{X_n - np}{\sqrt{npq}} < b\right] = \sum_{\left\{k \in \{0,1,\ldots,n\}: a < \frac{k-np}{\sqrt{npq}} < b\right\}} P\left[X_n = k\right]$$

$$= \sum_{\left\{k \in \{0,1,\ldots,n\}: a < x_{k,n} < b\right\}} \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x_{k,n}^2\right\}\left[1 + \alpha(n,k)\right]$$

Ahora bien, obsérvese que el conjunto $P = \{x_{k,n} : k \in \{0,\dots,n\}, a < x_{k,n} \leq b\} \cup \{a,b\}$ forma una partición del intervalo $[a,b]$, de manera que si $P = \{y_0, y_1, \dots, y_m\}$, con $a = y_0 < y_1 < \cdots < y_m = b$, entonces $\|P\| = \text{máx}\{y_j - y_{j-1} : j \in \{1,\dots,m\}\} = \frac{1}{\sqrt{npq}}$ y:

$$(y_1 - a)\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}a^2\right\} + \sum_{\left\{k\in\{0,1,\dots,n\}:a<x_{k,n}<b\right\}} \frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}x_{k,n}^2\right\}$$

$$-\frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}y_{m-1}^2\right\} + (b - y_{m-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}y_{m-1}^2\right\}$$

$$= \sum_{j=1}^m (y_j - y_{j-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}y_{j-1}^2\right\}$$

constituye una suma de Riemann de la función $\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x^2\right\}$, correspondiente a la partición P. Además:

$$0 \leq (y_1 - a)\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}a^2\right\} \leq \frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}$$

$$0 \leq \left(\frac{1}{\sqrt{npq}} - b + y_{m-1}\right)\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}y_{m-1}^2\right\} \leq \frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}$$

Por lo tanto:

$$\text{lím}_{n\rightsquigarrow\infty} \sum_{\left\{k\in\{0,1,\dots,n\}:a<x_{k,n}<b\right\}} \frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}x_{k,n}^2\right\}$$

$$= \text{lím}_{n\rightsquigarrow\infty} \sum_{j=1}^m (y_j - y_{j-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}y_{j-1}^2\right\}$$

$$-\text{lím}_{n\rightsquigarrow\infty}(y_1 - a)\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}a^2\right\} + \text{lím}_{n\rightsquigarrow\infty}\left(\frac{1}{\sqrt{npq}} - b + y_{m-1}\right)\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}y_{m-1}^2\right\}$$

$$= \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx$$

Por otra parte, de acuerdo con el teorema local de de Moivre-Laplace, $\text{lím}_{n\rightsquigarrow\infty}\alpha(n,k) = 0$ uniformemente en k, es decir, dada $\varepsilon > 0$, existe N tal que si $n \geq N$ y k es un entero no negativo tal que $np + a\sqrt{npq} \leq k \leq np + b\sqrt{npq}$, entonces $|\alpha(n,k)| < \varepsilon$. Así que, para $n \geq N$, se tiene:

$$\left|\sum_{\left\{k\in\{0,\dots,n\}:a<x_{k,n}<b\right\}} \frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}x_{k,n}^2\right\}\alpha(n,k)\right|$$

$$\leq \sum_{\left\{k\in\{0,\dots,n\}:a<x_{k,n}<b\right\}} \frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}x_{k,n}^2\right\}|\alpha(n,k)|$$

$$< \varepsilon \sum_{\left\{k\in\{0,\dots,n\}:a<x_{k,n}<b\right\}} \frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}x_{k,n}^2\right\}$$

Por lo tanto, para cualquier $\varepsilon > 0$, se tiene:

$\lim_{n\rightsquigarrow\infty}\left|\sum_{\left\{k\in\{0,\ldots,n\}:a<x_{k,n}<b\right\}}\frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x_{k,n}^2\right\}\alpha(n,k)\right|$

$\leq\varepsilon\lim_{n\rightsquigarrow\infty}\sum_{\left\{k\in\{0,\ldots,n\}:a<x_{k,n}<b\right\}}\frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x_{k,n}^2\right\}=\frac{\varepsilon}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx<\varepsilon$

Así que:

$\lim_{n\rightsquigarrow\infty}\sum_{\left\{k\in\{0,\ldots,n\}:a<x_{k,n}<b\right\}}\frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x_{k,n}^2\right\}\alpha(n,k)=0$

Por lo tanto:

$\lim_{n\rightsquigarrow\infty}P\left[a<\frac{X_n-np}{\sqrt{npq}}<b\right]$

$=\lim_{n\rightsquigarrow\infty}\sum_{\left\{k\in\{0,\ldots,n\}:a<x_{k,n}<b\right\}}\frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x_{k,n}^2\right\}[1+\alpha(n,k)]$

$=\lim_{n\rightsquigarrow\infty}\sum_{\left\{k\in\{0,1,\ldots,n\}:a<x_{k,n}<b\right\}}\frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x_{k,n}^2\right\}$

$+\lim_{n\rightsquigarrow\infty}\sum_{\left\{k\in\{0,1,\ldots,n\}:a<x_{k,n}<b\right\}}\frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x_{k,n}^2\right\}\alpha(n,k)$

$=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx$

■

Corolario 8.12. *Sea a cualquier número real y, para cada $n\in\mathbb{N}$, sea X_n una variable aleatoria con distribución binomial de parámetros n y $p\in(0,1)$. Entonces:*

$$\lim_{n\rightsquigarrow\infty}P\left[\frac{X_n-np}{\sqrt{npq}}>a\right]=\frac{1}{\sqrt{2\pi}}\int_a^\infty e^{-\frac{1}{2}x^2}dx$$

$$\lim_{n\rightsquigarrow\infty}P\left[\frac{X_n-np}{\sqrt{npq}}<a\right]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^a e^{-\frac{1}{2}x^2}dx$$

Demostración

Demostraremos primero que $\lim_{n\rightsquigarrow\infty}P\left[\frac{X_n-np}{\sqrt{npq}}>a\right]=\frac{1}{\sqrt{2\pi}}\int_a^\infty e^{-\frac{1}{2}x^2}dx$.

Para esto, como $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2}x^2}dx=1$, dada $\varepsilon>0$ se puede elegir $M>0$ suficientemente grande tal que $\frac{1}{\sqrt{2\pi}}\int_{-M}^M e^{-\frac{1}{2}x^2}dx>1-\frac{\varepsilon}{4}$. De esta forma se tiene $\frac{1}{\sqrt{2\pi}}\int_M^\infty e^{-\frac{1}{2}x^2}dx<\frac{\varepsilon}{4}$.

Sea ahora N_1 tal que, para cualquier $n\geq N_1$, se tenga:

$\left|P\left[-M<\frac{X_n-np}{\sqrt{npq}}<M\right]-\frac{1}{\sqrt{2\pi}}\int_{-M}^M e^{-\frac{1}{2}x^2}dx\right|<\frac{\varepsilon}{4}$

Entonces:

$P\left[-M < \frac{X_n - np}{\sqrt{npq}} < M\right] > \frac{1}{\sqrt{2\pi}} \int_{-M}^{M} e^{-\frac{1}{2}x^2} dx - \frac{\varepsilon}{4} > 1 - \frac{\varepsilon}{2}$

Así que:

$P\left[\frac{X_n - np}{\sqrt{npq}} \geq M\right] < \frac{\varepsilon}{2}$

Si $a \geq M$, entonces, para $n \geq N_1$, se tiene:

$P\left[\frac{X_n - np}{\sqrt{npq}} > a\right] \leq P\left[\frac{X_n - np}{\sqrt{npq}} \geq M\right] < \frac{\varepsilon}{2}$

y:

$\frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-\frac{1}{2}x^2} dx \leq \frac{1}{\sqrt{2\pi}} \int_{M}^{\infty} e^{-\frac{1}{2}x^2} dx < \frac{\varepsilon}{4}$

así que:

$\left|P\left[\frac{X_n - np}{\sqrt{npq}} > a\right] - \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-\frac{1}{2}x^2} dx\right| < \frac{\varepsilon}{2} < \varepsilon$

Si $a < M$, sea N_2 tal que, para cualquier $n \geq N_2$, se tenga:

$\left|P\left[a < \frac{X_n - np}{\sqrt{npq}} < M\right] - \frac{1}{\sqrt{2\pi}} \int_{a}^{M} e^{-\frac{1}{2}x^2} dx\right| < \frac{\varepsilon}{4}$

Entonces, para $n \geq \max\{N_1, N_2\}$, se tiene:

$\left|P\left[\frac{X_n - np}{\sqrt{npq}} > a\right] - \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-\frac{1}{2}x^2} dx\right|$

$= \left|P\left[a < \frac{X_n - np}{\sqrt{npq}} < M\right] + P\left[\frac{X_n - np}{\sqrt{npq}} \geq M\right] - \frac{1}{\sqrt{2\pi}} \int_{a}^{M} e^{-\frac{1}{2}x^2} dx - \frac{1}{\sqrt{2\pi}} \int_{M}^{\infty} e^{-\frac{1}{2}x^2} dx\right|$

$\leq \left|P\left[a < \frac{X_n - np}{\sqrt{npq}} < M\right] - \frac{1}{\sqrt{2\pi}} \int_{a}^{M} e^{-\frac{1}{2}x^2} dx\right| + P\left[\frac{X_n - np}{\sqrt{npq}} \geq M\right] + \frac{1}{\sqrt{2\pi}} \int_{M}^{\infty} e^{-\frac{1}{2}x^2} dx$

$< \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} = \varepsilon$

El otro límite se puede ahora obtener de la siguiente forma:

$\lim_{n \rightsquigarrow \infty} P\left[\frac{X_n - np}{\sqrt{npq}} < a\right] = 1 - \lim_{n \rightsquigarrow \infty} P\left[\frac{X_n - np}{\sqrt{npq}} > a\right] - \lim_{n \rightsquigarrow \infty} P\left[\frac{X_n - np}{\sqrt{npq}} = a\right]$

$= 1 - \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}x^2} dx.$

∎

Ejemplo 8.13. *Un productor de focos sabe que el 5% de su producción está defectuosa. El productor da una garantía en su envío de 10,000 focos, prometiendo la devolución del costo de la mercancía si más de α focos en el envío resultan defectuosos. ¿Qué* *tan pequeño puede escoger el productor el valor de α de tal manera que pueda confiar en que no tendrá que dar un reembolso en más del 1% de los envíos?*

Solución

Sea X el número de artículos defectuosos en un envío de 10,000 focos. X tiene entonces distribución binomial con parámetros $n = 10,000$ y $p = 0.05$. Por otra parte, se está buscando α tal que $P[X > \alpha] \leq 0.01$.

Utilizando el teorema de de Moivre-Laplace, se tiene:

$$P[X > \alpha] \approx P\left[Z \geq \frac{\alpha - np}{\sqrt{np(1-p)}}\right] = P\left[Z \geq \frac{\alpha - 500}{21.794}\right]$$

en donde Z es una variable aleatoria con distribución normal estándar.

Pero $P[Z \geq 2.33] \approx 0.01$, de manera que se busca α tal que $\frac{\alpha-500}{21.794} \geq 2.33$. Es decir, $\alpha \geq 500 + (21.794)(2.33) = 550.78$.

Por lo tanto, escogiendo $\alpha \geq 551$, el productor puede confiar en que no tendrá que dar un reembolso en más del 1% de los envíos.

Ejemplo 8.14 (**Movimiento browniano**). *Consideremos una partícula que se mueve a saltos sobre la recta real de tal manera que, comenzando en el origen, en cada intervalo de tiempo h, la partícula salta una distancia d hacia la derecha con probabilidad $\frac{1}{2}$ o una distancia d hacia la izquierda, también con probabilidad $\frac{1}{2}$. Llamemos S_n a la posición de la partícula en el tiempo nh.*

De acuerdo con lo expuesto en la sección 7.7, si para cada $n \in N$, definimos la variable aleatoria X_n como el número de saltos que da la partícula hacia la derecha en el tiempo nh, entonces la distribución de X_n es binomial con parámetros n y $p = \frac{1}{2}$ y, además, $S_n = d(2X_n - n)$.

Por el teorema de de Moivre-Laplace, tenemos:

$$\lim_{n \rightsquigarrow \infty} P\left[a < \frac{X_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx$$

Pero $\frac{X_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = \frac{S_n}{d\sqrt{n}}$, así que:

$$\lim_{n \rightsquigarrow \infty} P\left[a < \frac{S_n}{d\sqrt{n}} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx$$

Es decir, cuando n es grande, la distribución de S_n es aproximadamente normal con parámetros $\mu = 0$ y $\sigma^2 = d^2 n$.

Una caminata aleatoria de este tipo se utiliza para modelar el movimiento browniano, el cual consiste en el movimiento de una pequeña partícula que se encuentra suspendida en agua, de tal manera que los choques de las moléculas del agua con la partícula provocan que esta última se desplace sobre el líquido. Este tipo de movimiento fue observado y estudiado por Robert Brown, de lo cual se deriva su nombre. Más tarde, Albert Einstein lo explicó utilizando la teoría molecular y finalmente Norbert Wiener estableció el modelo probabilístico que permite estudiarlo.

Visto el movimiento browniano en una dimensión, digamos proyectado sobre el eje horizontal, se observa que la partícula se mueve avanzando y retrocediendo en una forma que parece azarosa, de ahí que se modele considerando primero una partícula que se mueve a saltos sobre la recta real de tal manera que, comenzando en un punto x*, en cada intervalo de tiempo* h*, la partícula salta una distancia* d *hacia la derecha con probabilidad* $\frac{1}{2}$ *o una distancia* d *hacia la izquierda, también con probabilidad* $\frac{1}{2}$*; después se hace tender* h *a cero, asumiendo además que* d^2 *es proporcional a* h*, lo cual es una hipótesis que viene del análisis físico.*

Llamemos S_t *a la posición de la partícula en el tiempo* $t > 0$ *y, considerando primero el movimiento a saltos, definamos* X_t *como el número de saltos que da la partícula hacia la derecha en el tiempo* t*. Sabemos entonces que si* n *es el más grande entero menor o igual que* $\frac{t}{h}$*, la distribución de* X_t *es binomial con parámetros* n *y* $p = \frac{1}{2}$ *y, además,* $S_t = x + d(2X_t - n)$*.*

Por el teorema de de Moivre-Laplace, tenemos:

$$\lim_{h \rightsquigarrow 0} P\left[a < \frac{X_t - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx$$

Pero, asumiendo $d^2 = h$ *(lo cual es posible hacer tomando las unidades adecuadas) y tomando* h *de la forma* $\frac{t}{m}$ *con* $m \in \mathbb{N}$*, se tiene* $\frac{X_t - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = \frac{S_t - x}{d\sqrt{n}} = \frac{S_t - x}{\sqrt{nh}} = \frac{S_t - x}{\sqrt{t}}$*, así que:*

$$\lim_{m \rightsquigarrow \infty} P\left[a < \frac{S_t - x}{\sqrt{t}} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx$$

Es decir, la distribución de S_t *es normal con parámetros* $\mu = 0$ *y* $\sigma^2 = t$*.*

La siguiente figura ilustra una trayectoria browniana:

▲

Para lograr una mejor aproximación de la distribución binomial mediante una distribución normal se requiere realizar una pequeña corrección antes de aplicar el teorema de de Moivre-Laplace. Ésta tiene que ver con el hecho de que se está aproximando una distribución discreta mediante una continua. Por ejemplo, si X es una variable aleatoria con distribución binomial de parámetros $n = 500$ y $p = 0.3$ y queremos encontrar un valor estimado de $P[X > 130]$, el teorema de de Moivre-Laplace nos permite escribir:

$$P[X > 130] = P\left[\frac{X_n - np}{\sqrt{npq}} > \frac{130-150}{\sqrt{105}}\right] = P\left[\frac{X_n - np}{\sqrt{npq}} > -1.9518\right]$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{-1.9518}^{\infty} e^{-\frac{1}{2}x^2} dx = 0.97452$$

Es decir, estamos aproximando la probabilidad $P[X > 130]$, que se obtiene como una suma de probabilidades, mediante el área de de una región bajo la gráfica de la función de densidad normal estándar. El valor exacto de $P[X > 130]$ se puede obtener también como un área bajo la gráfica de una función f definiendo ésta como sigue:

$$f(x) = \begin{cases} P[X = k] & \text{si } x \in (k - 0.5, k + 0.5] \text{ y } k \in \{0, \ldots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

con lo cual se tiene:

$$P[X > 130] = \int_{130.5}^{500.5} f(x)dx$$

Al aplicar el teorema de de Moivre-Laplace, lo que estamos haciendo es aproximar la función f mediante el Corolario 8.9, pero entonces obtendríamos una mejor aproximación considerando el intervalo de integración con la corrección de 0.5 que interviene en la definición de la función f. De esta forma, una mejor aproximación para $P[X > 130]$ se obtiene de la siguiente manera:

$P[X > 130] \approx P\left[Z > \frac{130.5-150}{\sqrt{105}}\right] = P[Z > -1.903] = \frac{1}{\sqrt{2\pi}}\int_{-1.903}^{\infty} e^{-\frac{1}{2}x^2}dx = 0.97148$

en donde Z es una variable aleatoria con distribución normal estándar.

El valor exacto de la probabilidad en consideración está dado por:

$P[X > 130] = \sum_{k=131}^{500}\binom{500}{k}(0.3)^k(0.7)^{500-k} = 0.97267$

Ejemplo 8.15. *Sea X una variable aleatoria con distribución binomial de parámetros $n = 5000$ y $p = 0.2$. Utilice el teorema de de Moivre Laplace para estimar $P[990 \leq X \leq 1020]$.*

Solución

$P[990 \leq X \leq 1020] \approx P\left[\frac{989.5-1000}{\sqrt{800}} \leq Z \leq \frac{1020.5-1000}{\sqrt{800}}\right] = P[-0.37123 \leq Z \leq 0.72478]$

$= \frac{1}{\sqrt{2\pi}}\int_{-0.37123}^{0.72478} e^{-x^2/2}dx = 0.41047$

en donde Z es una variable aleatoria con distribución normal estándar.

Ejemplo 8.16. *Utilice el teorema de de Moivre-Laplace para estimar la probabilidad de que en $10,000$ dígitos, seleccionados al azar, el número 7 no aparezca más de 985 veces.*

Solución

Sea X el número de veces que aparece el 7. X tiene entonces distribución binomial con parámetros $n = 10,000$ y $p = 0.1$. Utilizando el teorema de de Moivre-Laplace, se tiene:

$P[X \leq 985] \approx P\left[Z \leq \frac{985.5-1000}{\sqrt{900}}\right] = P[Z \leq -0.48333]$

$= \frac{1}{\sqrt{2\pi}}\int_{-0.48333}^{\infty} e^{-x^2/2}dx = 0.68557$

en donde Z es una variable aleatoria con distribución normal estándar.

8.4. Distribución exponencial

Definición 8.17 (**Distribución exponencial**). *Si λ es un número real positivo, se dice que una variable aleatoria X tiene distribución exponencial con parámetro λ si la función:*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad de X.

Una distribución de este tipo se obtiene al considerar el tiempo que transcurre entre dos ocurrencias consecutivas de eventos que ocurren aleatoriamente en el tiempo. Este esquema ya lo consideramos en la sección 7.4 al estudiar la distribución Poisson. Ahí consideramos un experimento aleatorio consistente en la observación de eventos que ocurren aleatoriamente en el tiempo de tal manera que si, para $t \geq 0$, X_t es el número de veces que ocurre el evento hasta el tiempo t, entonces la familia de variables aleatorias $\{X_t\}_{t\geq 0}$ forma un proceso de Poisson de parámetro λ. En particular, $X_0 = 0$ y si $s < t$, entonces la variable aleatoria $X_t - X_s$ tiene distribución Poisson con parámetro $\lambda(t-s)$.

Sea ahora T_1 el tiempo que transcurre desde el instante inicial $t = 0$ hasta que ocurre el primer evento. Obsérvese entonces que, para $t > 0$, el evento $[T_1 > t]$ ocurre si y sólo si ocurre el evento $[N_t = 0]$. Por lo tanto:

$P\left[T_1 > t\right] = P_0(t) = e^{-\lambda t}$

Así que:

$F_{T_1}(t) = 1 - e^{-\lambda t}$

Y entonces, una función de densidad de T_1 está dada por:

$$f_{T_1}(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Es decir, T_1 tiene distribución exponencial con parámetro λ.

Sea ahora $s > 0$ y T el tiempo que transcurre desde el instante s hasta que ocurre un evento más. Obsérvese entonces que, para $t > 0$, el evento $[T > t]$ ocurre si y sólo si no hay ocurrencias de eventos en el intervalo de tiempo $(s, s+t]$. Por lo tanto:

$P\left[T > t\right] = P\left[N_{s+t} - N_s = 0\right] = e^{-\lambda t}$

$P\left[T > t\right] = P\left[N_{(s,s+t]} = 0\right] = P\left[N_t = 0\right] = P_0(t) = e^{-\lambda t}$

Así que, nuevamente, T tiene distribución exponencial con parámetro λ.

A continuación se presentan las gráficas de algunas funciones de densidad tipo exponencial.

$\lambda = \frac{1}{8}$ $\lambda = 2$ $\lambda = 10$

Ejemplo 8.18. *Sea T una variable aleatoria con distribución exponencial de parámetro λ. Demuestre que, para cualquier par de números reales positivos s y t, se tiene:*

$$P\left[T > t + s \mid T > t\right] = P\left[T > s\right]$$

Supongamos ahora que el tiempo, en años, que un radio funciona, tiene distribución exponencial con parámetro $\lambda = \frac{1}{8}$. Si una persona compra un radio usado, ¿cuál es la probabilidad de que ese radio funcionará durante 8 años más?

Solución

$P\left[T > t + s \mid T > t\right] = \frac{P[T>t+s,T>t]}{P[T>t]} = \frac{P[T>t+s]}{P[T>t]} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P\left[T > s\right]$

Por otra parte, se busca $P\left[T > t + 8 \mid T > t\right]$, en donde t es el tiempo que ha funcionado el radio en el momento de la compra. Por la relación demostrada anteriormente, se tiene:

$P\left[T > t + 8 \mid T > t\right] = P\left[T > 8\right] = e^{-8\lambda}$

Así que la probabilidad de que el radio comprado funcionará durante 8 años más es igual a $e^{-1} = 0.368$.

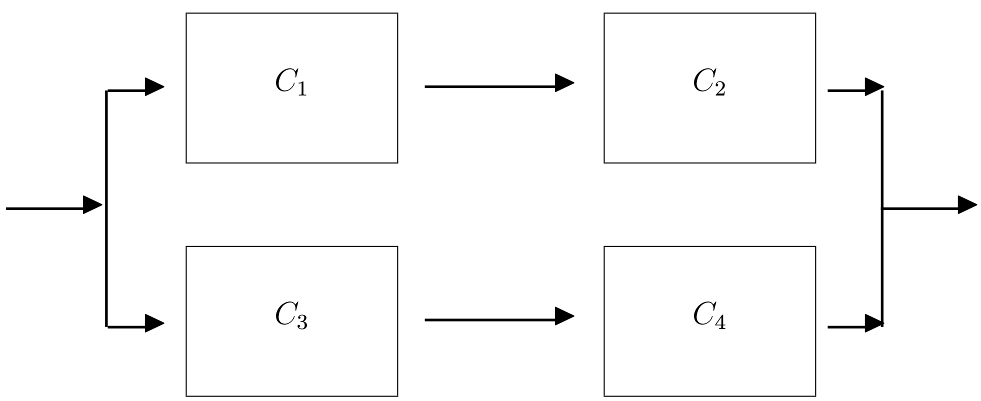
Ejemplo 8.19. *Sean T_1 y T_2 dos variables aleatorias independientes, ambas con distribución exponencial, T_1 con parámetro λ_1 y T_2 con parámetro λ_2. Demuestre que $T = \text{mín}(T_1, T_2)$ tiene distribución exponencial con parámetro $\lambda = \lambda_1 + \lambda_2$.*

Solución

$P\left[T > t\right] = P\left[\text{mín}(T_1, T_2) > t\right] = P\left[T_1 > t, T_2 > t\right]$

$= P\left[T_1 > t\right] P\left[T_2 > t\right] = e^{-\lambda_1 t} e^{-\lambda_2 t} = e^{-(\lambda_1+\lambda_2)t}$

Ejemplo 8.20. *La siguiente figura muestra 4 componentes, cada uno de los cuales tiene un tiempo de vida, independiente de los demás, distribuido exponencialmente con* *parámetro λ. Los 4 componentes están articulados formando un sistema de tal forma que, al funcionar el sistema, se activan los 4 componentes. El sistema permanece funcionando mientras una de las dos partes, la superior o la inferior, funcionan. A su vez cada una de éstas funciona cuando sus dos componentes funcionan. Encuentre la distribución del tiempo de vida del sistema.*



Solución

Sean T_1, T_2, T_3, T_4 los tiempos de vida de los componentes C_1, C_2, C_3, C_4, respectivamente, T el tiempo de vida del sistema, $X = \text{mín}\{T_1, T_2\}$ y $Y = \text{mín}\{T_3, T_4\}$. Entonces X y Y son independientes y ambas tienen distribución exponencial con parámetro 2λ. Así que, para $t > 0$, se tiene:

$$P[T \leq t] = P[\text{máx}\{X, Y\} \leq t] = P[X \leq t, Y \leq t]$$

$$= P[X \leq t] P[Y \leq t] = \left(1 - e^{-2\lambda t}\right)^2$$

Así que, una función de densidad de T está dada por:

$$f_T(t) = \begin{cases} 4\lambda e^{-2\lambda t}\left(1 - e^{-2\lambda t}\right) & \textit{si } t > 0 \\ 0 & \textit{en otro caso} \end{cases}$$

La gráfica de f_T se muestra a continuación para el caso $\lambda = 2$.

La siguiente figura muestra la gráfica de la función $t \mapsto P[T > t]$ *en línea sólida, mientras que la línea punteada muestra la gráfica de la función* $t \mapsto P[X > t] = P[Y > t]$*, ambas tomando* $\lambda = 2$*.*

8.5. Distribución gama

DEFINICIÓN 8.21 (**Distribución gama**). *Si* λ *es un número real positivo y* n *un número entero positivo, se dice que una variable aleatoria* X *tiene distribución gama con parámetros* n *y* λ *si la función:*

$$f_X(x) = \begin{cases} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad de X.

Al igual que la distribución exponencial, una distribución de este tipo se obtiene al considerar el tiempo que transcurre entre cierto número de ocurrencias de eventos que ocurren aleatoriamente en el tiempo.

Consideremos, como en la sección anterior, un experimento aleatorio consistente en la observación de eventos que ocurren aleatoriamente en el tiempo de tal manera que si, para $t \geq 0$, X_t es el número de veces que ocurre el evento hasta el tiempo t, entonces la familia de variables aleatorias $\{X_t\}_{t\geq 0}$ forma un proceso de Poisson de parámetro λ.

Sabemos entonces que, para k entero no negativo y $t > 0$:

$P_k(t) = P[N_t = k] = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$

Sea ahora X el tiempo que transcurre desde el origen hasta la r-sima ocurrencia del evento. Para $x > 0$, se tiene:

$P\left[X > x\right] = P\left[\text{ocurren menos de } r \text{ eventos en el intervalo de tiempo } [0, x]\right]$

$= \sum_{k=0}^{r-1} P[N_x = k] = \sum_{k=0}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!}$

Así que: $F_X(x) = 1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!}$

Por lo tanto, una función de densidad de X está dada por:

$f_X(x) = F_X'(x) = \lambda e^{-\lambda x} + \lambda \sum_{k=1}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!} - \lambda \sum_{k=1}^{r-1} \frac{k e^{-\lambda x}(\lambda x)^{k-1}}{k!}$

$= \lambda e^{-\lambda x} + \lambda \sum_{k=1}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!} - \lambda \sum_{k=1}^{r-1} \frac{e^{-\lambda x}(\lambda x)^{k-1}}{(k-1)!}$

$= \lambda \sum_{k=1}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!} - \lambda \sum_{k=1}^{r-2} \frac{e^{-\lambda x}(\lambda x)^k}{k!} = \lambda \frac{e^{-\lambda x}(\lambda x)^{r-1}}{(r-1)!} = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}$

Es decir, X tiene distribución gama con parámetros $\alpha = r$ y λ.

Ejemplo 8.22. *En una estación meteorológica se cuenta con un tipo de aparato de medición, el cual tiene un tiempo de vida que se distribuye exponencialmente, de tal manera que su tiempo promedio de vida es de* $1,000$ *horas. Si se utilizan* 10 *de esos aparatos en forma consecutiva, uno de ellos después de que el anterior ya no funciona, a) ¿cuál es la probabilidad de que alguno de los aparatos estará funcionando después de* $10,000$ *horas?, b) si después de* $5,000$ *horas todavía está funcionando alguno de los aparatos, ¿cuál es la probabilidad de que alguno de ellos siga funcionando después de* $10,000$ *horas más?*

Solución

Sea X el tiempo total de vida de los 10 aparatos, utilizados, como se indica, uno después del otro. X tiene entonces distribución gama con parámetros $\lambda = 0.001$ y $\alpha = 10$; por lo tanto:

a) $P\left[X > 10,000\right] = \int_{10,000}^{\infty} \frac{(0.001)^{10}}{9!} x^9 e^{-.001x} dx = 0.45793$

b) $P\left[X > 15,000 \mid X > 5000\right] = \frac{P[X>15,000]}{P[X>5,000]} = \frac{\int_{15,000}^{\infty} \frac{(0.001)^{10}}{9!} x^9 e^{-0.001x} dx}{\int_{5,000}^{\infty} \frac{(0.001)^{10}}{9!} x^9 e^{-0.001x} dx}$

$= \frac{0.0698537}{0.968172} = 0.0721501$

También se podría considerar la variable aleatoria X_t, igual al número de aparatos que han dejado de funcionar hasta el tiempo t y se tendría:

$P\left[X > t\right] = P\left[X_t < 10\right] = \sum_{k=0}^{9} \frac{(\lambda t)^k e^{-\lambda t}}{k!}$

Así que:

$P\left[X > 10,000\right] = e^{-10} \sum_{k=0}^{9} \frac{(10)^k}{k!} = 0.45793$

$P\left[X > 5,000\right] = e^{-5} \sum_{k=0}^{9} \frac{(5)^k}{k!} = 0.968172$

▲

A continuación se define la distribución gama de una manera más general.

Definición 8.23 (**Distribución gama**). *Si α y λ son números reales positivos, se dice que una variable aleatoria X tiene distribución gama con parámetros α y λ si la función:*

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad de X.

Verifiquemos que esta función así definida es una función de densidad:

$\int_0^\infty \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{1}{\lambda^\alpha} t^{\alpha-1} e^{-t} dt$

$= \frac{1}{\Gamma(\alpha)} \int_0^\infty t^{\alpha-1} e^{-t} dt = 1$

Aunque la obtención de una función de densidad tipo gama se remonta al año 1836, su importancia se dejó ver al obtenerse la función de densidad de la suma de los cuadrados de n variables aleatorias independientes, cada una de ella con distribución normal estándar, la cual es un caso particular de distribución gama. La obtención de tal función de densidad, en el caso general, se encuentra en un artículo de Bienaymé del año 1852.

EJEMPLO 8.24. *Sea X una variable aleatoria con distribución normal estándar, muestre que X^2 tiene distribución gama.*

Solución

Para $z > 0$, se tiene:

$F_{X^2}(z) = P\left[X^2 \leq z\right] = P\left[-\sqrt{z} \leq X \leq \sqrt{z}\right] = F_X(\sqrt{z}) - F_X(-\sqrt{z})$

Así que, una función de densidad de X^2 está dada por:

$f_{X^2}(z) = \frac{d}{dz}F_{X^2}(z) = \frac{1}{2\sqrt{z}}f_X(\sqrt{z}) + \frac{1}{2\sqrt{z}}f_X(-\sqrt{z}) = \frac{1}{\sqrt{z}}f_X(\sqrt{z})$

$= \frac{1}{\sqrt{z}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z} = \frac{1}{\sqrt{2\pi}}z^{-\frac{1}{2}}e^{-\frac{1}{2}z} = \frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}z^{-\frac{1}{2}}e^{-\frac{1}{2}z}}{\Gamma(\frac{1}{2})}$

Por lo tanto, X^2 tiene distribución gama con parámetros $\alpha = \frac{1}{2}$ y $\lambda = \frac{1}{2}$.

PROPOSICIÓN 8.25. *La función $f(x) = \frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}$ crece hasta alcanzar un máximo en $x = \frac{\alpha-1}{\lambda}$, después de lo cual decrece.*

Demostración

Sea $g(x) = x^{\alpha-1}e^{-\lambda x}$,entonces:

$g'(x) = (\alpha-1)x^{\alpha-2}e^{-\lambda x} - \lambda x^{\alpha-1}e^{-\lambda x} = x^{\alpha-2}e^{-\lambda x}(\alpha - 1 - \lambda x) = 0$

Así que g alcanza su valor máximo cuando $g'(x) = 0$, es decir, en $x = \frac{\alpha-1}{\lambda}$.

■

A continuación se presentan las gráficas de algunas funciones de densidad tipo gama.

$\alpha = \frac{5}{4},\ \lambda = 2$ $\alpha = 20,\ \lambda = 2$ $\alpha = 100,\ \lambda = 50$

Como lo mencionamos antes, un caso particular importante de distribución gama se obtiene al considerar la distribución de la suma de los cuadrados de n variables aleatorias independientes con distribución normal estándar. El resultado es una distribución gama con parámetros $\alpha = \frac{n}{2}$ y $\lambda = \frac{1}{2}$, la cual es conocida como distribución χ^2 con n grados de libertad.

Definición 8.26 (**Distribución $\boldsymbol{\chi^2}$**). *Sea $n \in \mathbb{N}$. Se dice que una variable aleatoria X tiene distribución χ^2 con n grados de libertad si su distribución es gama con parámetros $\alpha = \frac{n}{2}$ y $\lambda = \frac{1}{2}$, es decir, si la función:*

$$f_X(x) = \begin{cases} \frac{x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad de X.

A continuación se presentan las gráficas de algunas funciones de densidad tipo χ^2.

$n = 1$ $n = 40$

8.6. Distribuciones uniformes en el plano

De la misma manera en que, como vimos en la sección 8.1, una distribución uniforme continua en el intervalo (a, b) se obtiene al considerar el experimento aleatorio consistente en la elección al azar de un número real en (a, b) y definiendo X como el número que se obtiene, obtenemos una distribución uniforme en el plano al considerar el experimento aleatorio consistente en la elección al azar de un punto en una región R del plano con un área bien definida. Como en el caso de una dimensión, la elección al azar se interpreta en el sentido de que dos subconjuntos de R que se puedan sobreponer geométricamente uno sobre el otro deben de tener asignada la misma probabilidad. En esta situación, vimos en la sección 5.2 que la probabilidad de que el punto seleccionado pertenezca a la región S es igual al cociente del área de S entre el área de R.

Utilizando esta idea podemos obtener la distribución de una variable aleatoria que se genere en este contexto geométrico.

Ejemplo 8.27. *Un experimento aleatorio consiste en seleccionar al azar un punto (x, y) en el interior de un círculo de radio R. Sea X la distancia del punto seleccionado al centro del círculo. Encuentre una función de densidad de X.*

Solución

Para $x \in (0, R)$, el evento $[X \leq x]$ consiste del disco de radio x, por lo tanto:

$$F_X(x) = P[X \leq x] = \begin{cases} 0 & si\ x < 0 \\ \frac{\pi x^2}{\pi R^2} & si\ 0 \leq x < R \\ 1 & si\ x \geq R \end{cases}$$

Así que, una función de densidad de X está dada por:

$$f_X(x) = \begin{cases} \frac{2x}{R^2} & si\ 0 < x < R \\ 0 & en\ otro\ caso \end{cases}$$

Ejemplo 8.28. *Un experimento aleatorio consiste en seleccionar al azar un punto en el interior de un triángulo de base b y altura h. Sea X la distancia del punto seleccionado a la base. Encuentre una función de densidad de X.*

Solución

Para $x \in (0, h)$, el evento $[X \leq x]$ consiste del trapecio que se forma al intersectar el triángulo dado mediante la recta paralela a la base del triángulo y situada a una distancia x de la misma. El área de esta región es igual al área total del triángulo dado menos el área del triángulo de base la recta paralela mencionada.

Por lo tanto:

$$F_X(x) = P\left[X \leq x\right] = \begin{cases} 0 & si\ x \leq 0 \\ \frac{2}{bh}\left[\frac{1}{2}bh - \frac{1}{2h}b(h-x)^2\right] & si\ 0 < x < h \\ 1 & si\ x \geq h \end{cases}$$

$$= \begin{cases} 0 & si\ x \leq 0 \\ 1 - \frac{1}{h^2}(h-x)^2 & si\ 0 < x < h \\ 1 & si\ x \geq h \end{cases}$$

Así que, una función de densidad de X está dada por:

$$f_X(x) = \begin{cases} \frac{2}{h^2}(h-x) & si\ 0 < x < h \\ 0 & en\ otro\ caso \end{cases}$$

8.7. Distribución de funciones de variables aleatorias continuas

En ocasiones la distribución de una variable aleatoria X es conocida y se quiere obtener la distribución de una función de X. Ya hemos encontrado esta situación con anterioridad; en efecto, en el ejemplo 8.4 mostramos que si X es una variable aleatoria con distribución normal de parámetros μ y σ^2, entonces $Y = \frac{X-\mu}{\sigma}$ tiene distribución normal con parámetros 0 y 1 y en el ejemplo 8.24 mostramos que si X es una variable aleatoria con distribución normal estándar, entonces X^2 tiene distribución gama. Como se muestra en esos ejemplos, para encontrar la distribución de una función $f(X)$ de una variable aleatoria X absolutamente continua, se puede seguir el método siguiente: primero se expresa la función de distribución de $f(X)$ en términos de la función de distribución de X; después, la función de densidad de $f(X)$ se obtiene derivando la expresión anterior y utilizando el hecho de que la derivada de la función de distribución de X es igual a su función de densidad. En seguida se exponen más ejemplos de este tipo.

EJEMPLO 8.29. *Un experimento aleatorio consiste en seleccionar al azar un punto* (x, y) *en el interior de un círculo de radio* 1 *y centro en el origen. Después de eso, se traza un círculo* C *de centro en el origen que pasa por* (x, y)*. Encuentre la distribución del área de* C*.*

Solución

$$F_X(x) = P[X \leq x] = \begin{cases} 0 & si\ x < 0 \\ \frac{\pi x^2}{\pi R^2} & si\ 0 \leq x < R \\ 1 & si\ x \geq R \end{cases}$$

Sea r *el radio de* C*. De acuerdo con el ejercicio 8.27, una función de densidad de* r *está dada por:*

$$f_r(x) = \begin{cases} 2x & si\ 0 < x < 1 \\ 0 & en\ otro\ caso \end{cases}$$

Sea ahora A *el área de* C*, entonces, para* $y \in (0, \pi)$*, se tiene:*

$$F_A(y) = P[A \leq y] = P[\pi r^2 \leq y] = P\left[r \leq \sqrt{\tfrac{y}{\pi}}\right] = F_r\left(\sqrt{\tfrac{y}{\pi}}\right)$$

Por lo tanto, una función de densidad de A *está dada por:*

$$f_A(y) = \tfrac{d}{dy}F_A(y) = \tfrac{1}{2\sqrt{\pi y}} f_r\left(\sqrt{\tfrac{y}{\pi}}\right) = \tfrac{2}{2\sqrt{\pi y}}\sqrt{\tfrac{y}{\pi}} = \tfrac{1}{\pi}$$

Así que la distribución del área es uniforme en el intervalo $(0, \pi)$*.*

EJEMPLO 8.30. *En Mecánica Estadística, la velocidad* V *de una molécula de masa* m *en un gas, a temperatura* T*, se modela como una variable aleatoria con función de densidad dada por*

$$f_V(v) = \begin{cases} \sqrt{\frac{2m^3}{\pi k^3 T^3}}\, v^2 e^{-\frac{m}{2kT}v^2} & si\ v > 0 \\ 0 & en\ otro\ caso \end{cases}$$

en donde k *es la constante de Boltzmann. Encuentre una función de densidad de la energía* $E = \frac{1}{2}mV^2$*.*

Solución

Para $x > 0$*, se tiene:*

$$F_E(x) = P[E \leq x] = P\left[\tfrac{1}{2}mV^2 \leq x\right] = P\left[V \leq \sqrt{\tfrac{2x}{m}}\right] = F_V(\sqrt{\tfrac{2x}{m}})$$

Por lo tanto, una función de densidad de E *está dada por:*

$f_E(x) = \frac{d}{dx}F_E(x) = \sqrt{\frac{1}{2mx}}f_V(\sqrt{\frac{2x}{m}}) = \sqrt{\frac{1}{2mx}}\sqrt{\frac{2m^3}{\pi k^3T^3}}\frac{2x}{m}e^{-\frac{m}{2kT}\frac{2x}{m}} = \sqrt{\frac{1}{\pi k^3T^3}}x^{\frac{1}{2}}e^{-\frac{x}{kT}}$

Así que la distribución de la energía es gama con parámetros $\alpha = \frac{3}{2}$ *y* $\lambda = \frac{1}{kT}$.

Ejemplo 8.31. *Sea* X *una variable aleatoria distribuida uniformemente en el intervalo* $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. *Encuentre la función de distribución y una función de densidad de* $Y = \tan(X)$.

Solución

Para $y \in \mathbb{R}$, *se tiene:*

$P[Y \leq y] = P[\tan(X) \leq y] = P\left[-\frac{\pi}{2} \leq X \leq \arctan(y)\right] = \frac{1}{\pi}\left[\frac{\pi}{2} + \arctan(y)\right]$

Por lo tanto:

$F_Y(y) = \frac{1}{\pi}\left[\frac{\pi}{2} + \arctan(y)\right]$

Así que, una función de densidad de Y *está dada por:*

$f_Y(y) = \frac{1}{\pi(1+y^2)}$

A esta distribución se le conoce como ***distribución Cauchy***.

Ejemplo 8.32. *Sea* T *una variable aleatoria con distribución exponencial de parámetro* λ *y* $M > 0$. *Encuentre la distribución de* $X = \min\{T, M\}$.

Solución

$P[X > x] = P[\min\{T, M\} > x] = P[T > x, M > x]$

$$= \begin{cases} 1 & si\ x \leq 0 \\ P[T > x] & si\ x < M \\ 0 & si\ x \geq M \end{cases} = \begin{cases} 1 & si\ x \leq 0 \\ e^{-\lambda x} & si\ x < M \\ 0 & si\ x \geq M \end{cases}$$

Así que:

$$F_X(x) = P[X \leq x] = 1 - P[X > x] = \begin{cases} 0 & si\ x \leq 0 \\ 1 - e^{-\lambda x} & si\ 0 < x < M \\ 1 & si\ x \geq M \end{cases}$$

Obsérvese que la función de distribución de Z *es continua y estrictamente creciente en el intervalo* $(0, M)$, *pero tiene una discontinuidad en* $z = M$, *de manera que* Z *es una variable aleatoria que no es ni discreta ni continua.*

A continuación se muestra la gráfica de F_X *para el caso* $\lambda = \frac{1}{5}$ *y* $M = 5$.

$$\lambda = \frac{1}{5},\ M = 5$$

8.8. Simulación de distribuciones

Un problema importante en la Teoría de la Probabilidad consiste en la obtención de números que simulen los valores de una variable aleatoria con distribución dada. Por ejemplo, para generar números en el intervalo $[0,1)$ que simulen los valores de una variable aleatoria con distribución uniforme en ese intervalo, uno de los métodos más utilizados consiste en seleccionar un número natural M (llamado módulo y cuyo valor se toma grande, por ejemplo $M = 2^{32}$), un número natural $X_0 < M$ (llamado semilla) y dos números naturales más, a y c (que se toman pequeños, por ejemplo $a = 2$, $c = 1$). Se define entonces la sucesión $X_1, X_2, \ldots$, recursivamente, mediante la fórmula:

$$X_{n+1} = aX_n + c \bmod(M)$$

Los números $\frac{X_1}{M}, \frac{X_2}{M}, \ldots$ simulan entonces los valores de una variable aleatoria con distribución uniforme en el intervalo $[0,1)$.

Una vez resuelto el problema de simulación para la distribución uniforme, es posible resolverlo para cualquier otra distribución utilizando los resultados que se exponen en esta sección.

Proposición 8.33. *Sea X una variable aleatoria continua y F_X su función de distribución, entonces la variable aleatoria $Y = F_X(X)$ tiene distribución uniforme en el intervalo* $(0,1)$.

Demostración

Se tiene inmediatamente:

$$P[Y \leq y] = P[F_X(X) \leq y] = \begin{cases} 0 & \text{si } y < 0 \text{ ó } y = 0 \text{ y } F_X(x) > 0 \text{ para toda } x \\ 1 & \text{si } y \geq 1 \end{cases}$$

Por otra parte, si $y \in (0,1)$ ó $y = 0$ y $F_X(x) = 0$ para alguna x, definamos:

$x_0 = \sup\{x \in \mathbb{R} : F_X(x) = y\}$

Como F_X es continua, se tiene $F_X(x_0) = y$. Además, $F_X(x) > y$ para cualquier $x > x_0$ y $F_X(x) \leq y$ para cualquier $x < x_0$. Es decir, $F_X(x) \leq y$ si y sólo si $x \leq x_0$. Por lo tanto:

$P[F_X(X) \leq y] = P[X \leq x_0] = F_X(x_0) = y$

Así que:

$$P[Y \leq y] = P[F_X(X) \leq y] = \begin{cases} 0 & \text{si } y \leq 0 \\ y & \text{si } y \in (0,1) \\ 1 & \text{si } y \geq 1 \end{cases}$$

■

Cuando la función de distribución F_X tiene una inversa, F_X^{-1}, la proposición 8.33 permite obtener una variable aleatoria con función de distribución F_X a partir de una variable aleatoria Y con distribución uniforme en el intervalo $(0,1)$, simplemente definiendo $X = F_X^{-1}(Y)$.

EJEMPLO 8.34. *Sea X una variable aleatoria con distribución exponencial de parámetro λ, entonces:*

$$F_X(x) = \begin{cases} 0 & si\ x < 0 \\ 1 - e^{-\lambda x} & si\ x \geq 0 \end{cases}$$

Así que, $Y = 1 - e^{-\lambda X}$ tiene distribución uniforme en el intervalo $(0,1)$. Por lo tanto, si Y tiene distribución uniforme en el intervalo $(0,1)$, entonces $-\frac{1}{\lambda}\ln(1-Y)$ tiene distribución exponencial con parámetro λ.

Ahora bien, como la distribución de $1 - Y$ coincide con la de Y, se concluye que $-\frac{1}{\lambda}\ln Y$ tiene distribución exponencial con parámetro λ.

▲

Cuando la función de distribución F_X no es invertible, se requiere un resultado más fino para poder obtener una variable aleatoria con función de distribución F_X a partir de una variable aleatoria con distribución uniforme en el intervalo $(0,1)$. Este resultado se expone a continuación:

PROPOSICIÓN 8.35. *Sea X una variable aleatoria con función de distribución F_X, Y una variable aleatoria con distribución uniforme en el intervalo $(0,1)$ y definamos la función $d:(0,1) \mapsto \mathbb{R}$ mediante la relación $d(t) = \text{ínf}\{s \in \mathbb{R} : F_X(s) \geq t\}$, entonces la función de distribución de la variable aleatoria $d(Y)$ es F_X.*

Demostración

Obsérvese que d es una función no decreciente y, como F_X es continua por la derecha, se tiene $F_X(d(t)) \geq t$.

Vamos a demostrar que $[d(Y) \leq z] = [Y \leq F_X(z)]$. En efecto, si $d(Y(\omega)) \leq z$, entonces $Y(\omega) \leq F_X(d(Y(\omega))) \leq F_X(z)$ y si $Y(\omega) \leq F_X(z)$, entonces:

$$d(Y(\omega)) = \text{ínf}\{s \in \mathbb{R} : F_X(s) \geq Y(\omega)\} \leq z$$

Por lo tanto, se tiene:

$$P[d(Y) \leq z] = P[Y \leq F_X(z)] = F_X(z)$$

∎

COROLARIO 8.36. *Sea X una variable aleatoria con función de distribución F_X, Y una variable aleatoria con distribución uniforme en el intervalo $(0,1)$ y definamos la función $c:(0,1) \mapsto \mathbb{R}$ mediante la relación $c(t) = \text{ínf}\{s \in \mathbb{R} : F_X(s) > t\}$, entonces la función de distribución de la variable aleatoria $c(Y)$ es F_X.*

Demostración

Definamos $\mathcal{C}$ como el conjunto de puntos $t \in (0,1)$ para los cuales existe un intervalo de longitud positiva en el cual F_X es constante e igual a t.

Obsérvese que $c(t) = d(t)$ excepto en los puntos del conjunto $\mathcal{C}$. Además, d es discontinua en los puntos t que pertenecen $\mathcal{C}$. Por otra parte, por el lema 6.17, el conjunto de puntos en los cuales d es discontinua es finito o infinito numerable. Por lo tanto, $c(t) = d(t)$ excepto a lo más en un conjunto infinito numerable.

Finalmente $c(Y) = d(Y)$ excepto en el conjunto $[Y \in \mathcal{C}]$. Pero como $\mathcal{C}$ es a lo más un conjunto infinito numerable, se tiene $P[Y \in \mathcal{C}] = 0$. Por lo tanto, $P[c(Y) = d(Y)] = 1$. Así que:

$$P[c(Y) \leq z] = P[d(Y) \leq z] = F_X(z)$$

∎

EJEMPLO 8.37. *Sea X una variable aleatoria con distribución geométrica de parámetro p, entonces:*

$$F_X(x) = \begin{cases} 0 & si\ x < 0 \\ \sum_{k=0}^{[[x]]} p(1-p)^k & si\ x \geq 0 \end{cases} = \begin{cases} 0 & si\ x < 0 \\ 1-(1-p)^{[[x]]+1} & si\ x \geq 0 \end{cases}$$

en donde $[[x]]$ *es el mayor entero menor o igual a* x.

Así que, definiendo $\lambda = -\ln(1-p)$, *se tiene:*

$c(t) = \text{ínf}\{s \in \mathbb{R} : F_X(s) > t\} = \text{ínf}\left\{s \in \mathbb{R} : 1-(1-p)^{[[s]]+1} > t\right\}$

$= \left[\left[\text{ínf}\left\{s \in \mathbb{R} : 1-e^{-\lambda s} \geq t\right\}\right]\right] = \left[\left[\text{ínf}\left\{s \in \mathbb{R} : s \geq -\frac{1}{\lambda}\ln(1-t)\right\}\right]\right]$

$= \left[\left[-\frac{1}{\lambda}\ln(1-t)\right]\right]$

Por lo tanto, si U *tiene distribución uniforme en el intervalo* $(0,1)$, *entonces* $\left[\left[\frac{\ln(1-U)}{\ln(1-p)}\right]\right]$ *tiene distribución geométrica con parámetro* p.

Ahora bien, como U *y* $1-U$ *tienen la misma distribución, entonces* $\left[\left[\frac{\ln U}{\ln(1-p)}\right]\right]$ *tiene distribución geométrica con parámetro* p.

Obsérvese que, de acuerdo con el ejemplo 8.34, $\frac{\ln U}{\ln(1-p)} = -\frac{1}{\lambda}\ln U$ *tiene distribución exponencial con parámetro* λ, *de manera que, dado* $p \in (0,1)$, *si definimos* $\lambda = -\ln(1-p)$ *y* Z *es una variable aleatoria con distribución exponencial de parámetro* λ, *entonces* $[[Z]]$ *tiene distribución geométrica con parámetro* p.

Ejemplo 8.38. *Sea* X *una variable aleatoria con distribución Poisson de parámetro* λ, *entonces:*

$$F_X(x) = \begin{cases} 0 & si\ x < 0 \\ \sum_{k=0}^{[[x]]} \frac{\lambda^k e^{-\lambda}}{k!} & si\ x \geq 0 \end{cases}$$

en donde $[[x]]$ *es el mayor entero menor o igual a* x.

Así que:

$c(t) = \text{ínf}\{s \in \mathbb{R} : F_X(s) > t\} = \text{ínf}\left\{s \in \mathbb{R} : \sum_{k=0}^{[[s]]} \frac{\lambda^k e^{-\lambda}}{k!} > t\right\}$

$= \text{ínf}\left\{j \in \{0,1,\ldots\} : \sum_{k=0}^{j} \frac{\lambda^k e^{-\lambda}}{k!} > t\right\}$

Por lo tanto, si Y *tiene distribución uniforme en el intervalo* $(0,1)$, *entonces:*

$c(Y) = \text{ínf}\left\{j \in \{0,1,\ldots\} : \sum_{k=0}^{j} \frac{\lambda^k e^{-\lambda}}{k!} > Y\right\}$

tiene distribución Poisson con parámetro λ.

▲

La generación de una variable aleatoria con distribución Poisson puede obtenerse también a partir del hecho de que cuando se tiene un experimento aleatorio consistente en la observación de eventos que ocurren aleatoriamente en el tiempo de tal manera que si, para $t \geq 0$, X_t es el número de veces que ocurre el evento hasta el tiempo t, entonces la familia de variables aleatorias $\{X_t\}_{t\geq 0}$ forma un proceso de Poisson de parámetro λ. En esta situación, como vimos en la sección 8.4, los tiempos entre ocurrencias sucesivas tienen distribución exponencial con parámetro λ.

Dado $\lambda > 0$, consideremos entonces una sucesión de variables aleatorias independientes, T_1, T_2, ..., todas ellas con distribución exponencial de parámetro λ. La sucesión de variables aleatorias T_1, $T_1 + T_2$, $T_1 + T_2 + T_3$, ..., simula entonces los tiempos de ocurrencia de eventos que ocurren aleatoriamente en el tiempo. Llamemos X_t al número de eventos que ocurren hasta el tiempo t. Entonces, el evento $[X_t = 0]$ ocurre si y sólo si $T_1 > t$ y, para $k \in \mathbb{N}$, el evento $[X_t = k]$ ocurre si y sólo si $T_1 + \ldots + T_k \leq t$ y $T_1 + \ldots + T_k + T_{k+1} > t$. Es decir:

$$X_t = \text{mín}\left\{i \in \mathbb{N} \mid \textstyle\sum_{j=1}^{i} T_j > t\right\} - 1$$

es una variable aleatoria con distribución Poisson de parámetro λt.

De acuerdo con el ejemplo 8.34, cada variable aleatoria T_j puede generarse mediante una variable aleatoria U_j con distribución uniforme en el intervalo $(0, 1)$, definiendo $T_j = -\frac{1}{\lambda} \ln U_j$. De esta forma, si U_1, U_2, ... es una sucesión de variables aleatorias independientes, todas ellas con distribución uniforme en el intervalo $(0, 1)$, y λ en un número real positivo, entonces la variable aleatoria

$$X = \text{mín}\left\{i \in \mathbb{N} \mid \textstyle\sum_{j=1}^{i} -\frac{1}{\lambda} \ln U_j > 1\right\} - 1$$

tiene distribución Poisson con parámetro λ.

Finalmente, obsérvese que:

$$X = \text{mín}\left\{i \in \mathbb{N} \mid \textstyle\sum_{j=1}^{i} -\frac{1}{\lambda} \ln U_j > 1\right\} - 1 = \text{mín}\left\{i \in \mathbb{N} \mid -\frac{1}{\lambda} \ln \prod_{j=1}^{i} U_j > 1\right\} - 1$$

$$= \text{mín}\left\{i \in \mathbb{N} \mid \ln \textstyle\prod_{j=1}^{i} U_j < -\lambda\right\} - 1 = \text{mín}\left\{i \in \mathbb{N} \mid \prod_{j=1}^{i} U_j < e^{-\lambda}\right\} - 1$$

EJERCICIOS

Ejercicio 8.1. *En una parada de autobús, el tiempo de llegada de éste se distribuye uniformemente en el intervalo que va de las* $7:00$ *a las* $7:15$ *hrs. y el siguiente autobús pasa exactamente* 15 *minutos después del primero. Si el tiempo de llegada de una persona a esa parada se distribuye uniformemente en el intervalo que va de las* $7:10$ *a las* $7:15$ *hrs. Si* T *es el tiempo que espera la persona, desde que llega a la parada hasta que pasa un autobús, encuentre la distribución de* T.

Ejercicio 8.2. *Sea* X *una variable aleatoria con distribución uniforme en el intervalo* $(0,1)$. *Encuentre una función de densidad de* $Y = 1 - \sqrt{X}$.

Ejercicio 8.3. *Sea* X *una variable aleatoria con distribución uniforme en el intervalo* $(-1,1)$. *Encuentre una función de densidad de a)* $Y = |X|$ *y b)* $Z = X^2$.

Ejercicio 8.4. *Sea* X *una variable aleatoria distribuida uniformemente en el intervalo* $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. *Encuentre la función de distribución y una función de densidad de* $Y = \cos(X)$.

Ejercicio 8.5. *Sea* A *un evento que puede ocurrir en algún momento entre los tiempos* 0 *y* T *de tal manera que la probabilidad de que ocurra es igual a* p *y, en caso de que ocurra, si* X *es el tiempo en el cual ocurre, entonces* X *tiene distribución uniforme en el intervalo* $[0,T]$. *Sabiendo que el evento* A *no ha ocurrido hasta el tiempo* $t \in (0,T)$, *encuentre la probabilidad de que* A *ocurra en el intervalo* $(t,T]$.

Ejercicio 8.6. *Sea* X *una variable aleatoria con distribución normal de parámetros* μ *y* σ^2. *Encuentre* $P[|X-\mu| \leq \sigma]$.

Ejercicio 8.7. *Supongamos que el peso,* X, *de una persona, que se selecciona al azar de una determinada población, se distribuye normalmente con parámetros* μ *y* σ^2. *Si* $P[X \leq 70] = \frac{1}{2}$ *y* $P[X \leq 60] = \frac{1}{4}$, *encuentre* μ, σ *y* $P[X \geq 80]$. *¿Qué porcentaje de la gente de la población que pesa al menos 80 kilos, pesa más de 90 kilos?*

Ejercicio 8.8. *Sea* X *una variable aleatoria distribuida normalmente con parámetros* $\mu = 1$ *y* $\sigma^2 = 4$. *Encuentre* $P[5 \leq X^2 + 1 \leq 10]$.

Ejercicio 8.9. *Se ha determinado que en una determinada región la precipitación anual de lluvia tiene distribución normal. Si las estadísticas muestran que en el* 15% *de los casos la precipitación ha sido de más de* 45 *pulgadas y en el* 3% *ha sido de menos de* 30 *pulgadas, ¿cuál es la probabilidad de que en los próximos* 5 *años, por lo menos en uno de ellos la precipitación sea de más de* 50 *pulgadas?*

Ejercicio 8.10. *En una fábrica se producen componentes electrónicos cuyo voltaje tiene distribución normal de parámetros* μ *y* σ^2. *El departamento de control de calidad checa los componentes y todos aquellos con un voltaje menor o igual a* b *son* *rechazados. Encuentre la distribución del voltaje de los componentes que pasan el control de calidad.*

Ejercicio 8.11. *Sea X una variable aleatoria con distribución binomial de parámetros $n = 1,000$ y $p = 0.3$. Utilice el teorema de de Moivre-Laplace para estimar $P[X = 310]$. Encuentre además el valor exacto de esa probabilidad y compare los resultados. Con los mismos datos, estime $P[280 < X < 310]$.*

Ejercicio 8.12. *Supongamos que la probabilidad de que un recién nacido sea hombre es igual a 0.515. Estime la probabilidad de que entre 10 mil recién nacidos, haya más hombres que mujeres.*

Ejercicio 8.13. *Un cierto tipo de semilla tiene una probabilidad de germinar igual a 0.8. Estime la probabilidad de que, en un paquete de $1,000$ semillas, por lo menos el 65% germinará.*

Ejercicio 8.14. *Estime la probabilidad de que al lanzar 10,000 veces un dado se obtengan por lo menos 1680 1's.*

Ejercicio 8.15. *Estime el más pequeño número natural k tal que la probabilidad de que el número de soles que se obtienen en $1,000$ lanzamientos de una moneda esté comprendido entre 480 y k, inclusive, sea mayor que 0.6.*

Ejercicio 8.16. *Supongamos que el tiempo de vida, en horas, de un cierto producto tiene distribución exponencial con parámetro $\lambda = \frac{1}{50}$. Consideremos una población compuesta por 100 de esos productos y sea X el número de productos en la población que duran más de 50 horas. Estime $P[X > 40]$.*

Ejercicio 8.17. *Utilice el teorema de de Moivre-Laplace para estimar el más pequeño valor de n con la propiedad de que al lanzar n veces una moneda, la probabilidad de que el porcentaje de las veces en que se obtiene sol esté comprendido entre 49% y 51% sea mayor o igual a 0.95.*

Ejercicio 8.18. *Los focos producidos en una cierta fábrica tienen un tiempo de vida que se distribuye exponencialmente y se sabe que el 50% de los focos que se producen tienen un tiempo de vida mayor a $1,000$ horas. Estime la probabilidad de que entre $1,000$ focos, seleccionados al azar de la producción de esa fábrica, haya más de 275 con un tiempo de vida superior a $2,000$ horas.*

Ejercicio 8.19. *Sea X una variable aleatoria con distribución exponencial de parámetro λ y supongamos que $P[X \geq 10] = \frac{1}{2}$. Encuentre un número t tal que $P[X \geq t] = 0.9$.*

Ejercicio 8.20. *La vida de una partícula radioactiva se modela frecuentemente con una distribución exponencial y se define la vida media de la partícula como el tiempo que transcurre hasta que la radioactividad de la partícula se reduce a la mitad. Si la vida media del uranio es de 7.13×10^{-8} años, determine el tiempo que se requiere para que la radioactividad de una masa de uranio se reduzca un 90%.*

Ejercicio 8.21. *Supongamos que se tiene un número muy grande de partículas radioactivas, cada una de las cuales tiene un tiempo de estabilización que se distribuye exponencialmente con parámetro λ. Si la mitad de las partículas se estabiliza durante el primer segundo, ¿cuánto tiempo se requiere para que se estabilice el 75 % de ellas?*

Ejercicio 8.22. *Sea T una variable aleatoria absolutamente continua con función de densidad continua y tal que, para cualquier par de números reales positivos s y t, se tiene:*

$$P\left[T > t+s \mid T > t\right] = P\left[T > s\right]$$

Demuestre que T tiene distribución exponencial.

Ejercicio 8.23. *Una cierta máquina funciona bien únicamente si por lo menos 3 de sus 5 motores funcionan. Cada motor opera independientemente de los otros y el tiempo T que permanece funcionando tiene como función de densidad a $f(x) = xe^{-x}$ para $x > 0$. Asumiendo que los 5 motores se ponen a funcionar al mismo tiempo, encuentre la distribución del tiempo que la máquina funciona bien.*

Ejercicio 8.24. *El tiempo, en horas, que le toma a un técnico reparar una máquina tiene distribución exponencial con parámetro $\lambda = 2$. ¿Cuál es la probabilidad de que ese técnico pueda reparar 3 máquinas en menos de 2 horas?*

Ejercicio 8.25. *Sea $Y = (X-3)^2 + 1$, en donde X es una variable aleatoria con distribución gama de parámetros $\alpha = 2$ y λ. Encuentre $P\left[Y \geq 5\right]$.*

Ejercicio 8.26. *Sea X una variable aleatoria con distribución gama de parámetros α y λ. Encuentre la distribución de $Y = cX$, en donde c es una constante positiva.*

Ejercicio 8.27. *Supongamos que la vida en horas de cada lámpara, de una cierta clase, es una variable aleatoria T con función de densidad dada por:*

$$f_T(t) = \begin{cases} \frac{100}{t^2} & \text{si } t > 100 \\ 0 & \text{en otro caso} \end{cases}$$

¿Cuál es la probabilidad de que de 4 de esas lámparas, a) ninguna tenga que ser sustituida durante las primeras 150 horas de uso? y b) las 4 tengan que reemplazarse durante las primeras 150 horas de uso?

Ejercicio 8.28. *Un experimento aleatorio consiste en seleccionar, al azar, un punto al interior del triángulo formado por la recta $y = -x + 1$ y los ejes coordenados. Sea Y la distancia del punto seleccionado al eje y. Encuentre una función de densidad de Y.*

Ejercicio 8.29. *Un experimento aleatorio consiste en seleccionar, al azar, un punto sobre la base AB de un triángulo equilátero ABC cuyos lados miden 2 unidades cada* *uno. Sea X la distancia del punto seleccionado al vértice C. Encuentre una función de densidad de X.*

Ejercicio 8.30. *Un experimento aleatorio consiste en seleccionar al azar un punto (x, y) del cuadrado $0 \leq x \leq 1$, $0 \leq y \leq 1$. Sea X la distancia del punto seleccionado al origen. Encuentre una función de densidad de X.*

Ejercicio 8.31. *Un experimento aleatorio consiste en seleccionar al azar un punto (x, y) del cuadrado $0 \leq x \leq 1$, $0 \leq y \leq 1$. Sea Z la suma $x + y$. Encuentre una función de densidad de Z.*

Ejercicio 8.32. *Un experimento aleatorio consiste en seleccionar, al azar, un punto en el interior del cuadrado $0 \leq x \leq 1$, $0 \leq y \leq 1$. Sea X la distancia del punto seleccionado a la diagonal de pendiente 1. Encuentre una función de densidad de X.*

Ejercicio 8.33. *Un experimento aleatorio consiste en seleccionar al azar un punto en el interior del rombo de vértices $(0, 1)$, $(2, 0)$, $(0, -1)$ y $(-2, 0)$. Sea X la distancia del punto seleccionado a la diagonal menor del rombo. Encuentre una función de densidad de X.*

Ejercicio 8.34. *Dos personas A y B juegan el siguiente juego: cada persona elige al azar, de manera independiente de la otra, un punto sobre el intervalo $(0, 1)$. Si la distancia entre los dos puntos seleccionados es menor que una cantidad d, fija de antemano, el juego es ganado por A; de otra manera, el juego es ganado por B. Encuentre el valor de d para el cual ambas personas tienen la misma probabilidad de ganar el juego.*

Ejercicio 8.35. *Sea X una variable aleatoria continua con función de densidad f. Encuentre una función de densidad de $Y = aX + b$, en donde a y b son constantes y $a \neq 0$. Aplique el resultado al caso en que X tiene distribución uniforme en el intervalo $(0, 1)$.*

Ejercicio 8.36. *Sea X una variable aleatoria continua, no negativa, con función de densidad f. Encuentre una función de densidad de $Y = \sqrt{X}$. Aplique el resultado al caso en que X tiene distribución gama con parámetros α y λ.*

Ejercicio 8.37. *Sea X una variable aleatoria continua con función de densidad f. Encuentre una función de densidad de $Y = |X|$. Aplique el resultado al caso en que X tiene distribución a) normal con parámetros μ y σ^2 y b) uniforme en el intervalo $(-1, 2)$.*

Ejercicio 8.38. *Sea X una variable aleatoria continua con función de densidad f. Encuentre una función de densidad de $Y = e^X$. Aplique el resultado al caso en que X tiene distribución a) normal con parámetros μ y σ^2 y b) gama con parámetros α y λ.*

EJERCICIO 8.39. *Un función de densidad f, de una variable aleatoria continua X, está dada por la fórmula siguiente:*

$$f(x) = \begin{cases} cx^2 & si\ 0 < x < 1 \\ \frac{x}{2} & si\ 1 \leq x < 2 \\ 0 & en\ otro\ caso \end{cases}$$

en donde c es una constante. Encuentre una función de densidad de la variable aleatoria $Y = (1 - X)^2$.

EJERCICIO 8.40. *Muestre con un ejemplo que si X no es continua entonces la distribución de $F_X(X)$ no necesariamente es uniforme.*

EJERCICIO 8.41. *Sea X una variable aleatoria con función de distribución F_X, continua y estrictamente creciente, y definamos la función $d : (0,1) \mapsto \mathbb{R}$ mediante la relación:*

$$d(t) = \text{ínf}\,\{s \in \mathbb{R} : F_X(s) \geq t\}$$

Demuestre que d es la inversa de F_X.

EJERCICIO 8.42. *Demuestre directamente que si Y tiene distribución uniforme en el intervalo $(0,1)$, entonces, dado $\lambda > 0$, $-\frac{1}{\lambda}\ln Y$ tiene distribución exponencial con parámetro λ.*

EJERCICIO 8.43. *Demuestre directamente que si Y tiene distribución uniforme en el intervalo $(0,1)$, entonces, dado $p \in (0,1)$, $\left[\left[\frac{\ln Y}{\ln(1-p)}\right]\right]$ tiene distribución geométrica con parámetro p, en donde $[[x]]$ es el mayor entero menor o igual a x.*

EJERCICIO 8.44. *Demuestre directamente que si $p \in (0,1)$, $\lambda = -\ln(1-p)$ y Z es una variable aleatoria con distribución exponencial de parámetro λ, entonces, $[[Z]]$ tiene distribución geométrica con parámetro p, en donde $[[x]]$ es el mayor entero menor o igual a x.*

EJERCICIO 8.45. *Demuestre directamente que si Y tiene distribución uniforme en el intervalo $(0,1)$, entonces, dado $\lambda > 0$:*

$$c(Y) = \text{ínf}\left\{j \in \{0,1,\ldots\} : \sum_{k=0}^{j} \frac{\lambda^k e^{-\lambda}}{k!} > Y\right\}$$

tiene distribución Poisson con parámetro λ.

CAPÍTULO 9

ESPERANZAS

El principio fundamental de mi método consiste en que, en los juegos de azar, el valor de la suerte o de lo que espera cada jugador es precisamente lo que requeriría tener para que, jugando sin ventaja ni desventaja, tuviera una suerte o una espera similar. Por ejemplo, si alguien esconde tres monedas en una mano y siete en la otra y me deja escoger una de las dos alternativas, esta ventaja es para mí lo mismo que si me diera cinco monedas, porque con cinco monedas me encontraría en el caso de tener la misma espera para tres o para siete monedas en un juego sin ventaja ni desventaja.

Jacques Bernoulli

En este capítulo se estudia uno de los conceptos básicos de la Teoría de la Probabilidad, el de Esperanza de una variable aleatoria. Su importancia es comparable con la del concepto mismo de probabilidad de un evento. De hecho, el concepto de esperanza y el de probabilidad surgen en forma paralela cuando, a mediados del siglo $XVII$, se inicia la Teoría de la Probabilidad como disciplina matemática. Es Christiaan Huygens quien introdujo este concepto en su libro "Du calcul dans les jeux de hasard", publicado en el año 1657. En esa época, Blaise Pascal y Pierre de Fermat habían resuelto algunos problemas de probabilidad, con métodos que sentarían las bases para el desarrollo de una nueva disciplina matemática, el Cálculo de Probabilidades. Huygens resolvió, con sus propios métodos, los problemas que antes habían resuelto Pascal y Fermat y algunos otros más. Uno de los aspectos interesantes de la metodología utilizada por Huygens es que en ningún momento se consideran ahí probabilidades de eventos, todas las soluciones están basadas en el cálculo de esperanzas, lo cual hacía ver ya que este concepto podía tomarse como primario, previo incluso al de probabilidad de un evento, y a partir de él desarrollar la nueva disciplina. La historia no fue de ese modo pues el concepto que prevaleció como primario fue el de probabilidad. Sin embargo, la historia misma mostraría más adelante que esta dualidad de importancia, entre el concepto de esperanza y el de probabilidad, que se dio al inicio del desarrollo de la Teoría de la Probabilidad como disciplina matemática, tenía fuertes raíces pues al evolucionar el concepto de probabilidad, hasta fusionarse con el de medida en los primeros años de este siglo, resultó palpable la estrecha relación entre ambos conceptos, de tal manera que, efectivamente, cualquiera de los dos puede tomarse como punto de partida, quedando inmersos uno dentro del otro. Esto último ya no únicamente dentro del contexto de la Teoría de la Probabilidad, sino dentro del contexto más amplio de la Teoría de la Medida, en donde el concepto de probabilidad corresponde al de medida y el de esperanza al de integral.

Como motivación de la definición de la Esperanza, se expone a continuación su formulación basada en la idea de juego justo, siguiendo el trabajo original de Huygens.

Definición 9.1 (**Juego justo**). *Consideremos un juego con un número finito de participantes, de tal manera que uno de ellos resultará el ganador, determinándose quién, de manera específica, con base en el resultado de un experimento aleatorio $\mathcal{E}$. Diremos entonces que este juego es justo si puede ser considerado como combinación de juegos en cada uno de los cuales ocurre alguna de las siguientes dos situaciones:*

(i) *Si para un participante hay una probabilidad p de ganar una cantidad x, entonces para cualquier otro participante hay la misma probabilidad p de ganar x.*
(ii) *Si para un participante hay una probabilidad p de ganar una cantidad x, entonces hay la misma probabilidad p de perderla.*

Definición 9.2 (**Valor de un juego**). *Consideremos un juego en el cual participa una persona Q, de tal manera que, dependiendo del resultado de un experimento aleatorio $\mathcal{E}$, Q tiene la posibilidad de ganar algún valor. El valor del juego para Q es una cantidad x, la cual puede ser negativa, tal que existe un juego justo que le ofrece a Q las mismas posibilidades de ganar que el juego original y en el cual paga la cantidad x por participar*[1].

Ejemplo 9.3. *Sea $\mathcal{E}$ un experimento aleatorio cuyos posibles resultados son equiprobables, $A_1, \ldots, A_n$ una partición del correspondiente espacio muestral y $x_1, \ldots, x_n$ n números reales. Supongamos que una persona Q participa en un juego consistente en la realización del experimento aleatorio $\mathcal{E}$, de tal manera que si ocurre el evento A_i entonces recibe x_i pesos. ¿Cuál es el valor de este juego para Q?*

[1]Esta definición requeriría de la demostración de que tal número x, en caso de existir, es único. Sin embargo, como no se trata aquí de desarrollar una teoría del valor de juegos, no se abunda sobre este tema. La definición tiene por único objeto la motivación de la definición de esperanza que se da más adelante.

Solución

Sea r_i el número de elementos que tiene A_i y supongamos que hay n grupos de personas de tal manera que el i-ésimo grupo está formado por r_i personas y que Q forma parte del n-simo grupo. Consideremos entonces un juego entre esas $N = r_1 + \cdots + r_n$ personas, consistente en la realización del experimento $\mathcal{E}$, de tal manera que, en una relación uno a uno, se asocia con cada una de las personas del i-ésimo grupo un elemento del evento A_i. Se conviene que cada una de las personas paga una cierta cantidad x por participar en el juego y quien lo gane recibe toda la cantidad acumulada al inicio, es decir, Nx. Evidentemente, éste es un juego justo. Supongamos además que Q conviene con cada una de las personas del grupo i que en caso de que alguna de las dos gane el juego, el ganador dará al otro x_i pesos. Claramente, con esta nueva consideración, el juego sigue siendo justo. De acuerdo a lo establecido, en caso de que Q gane el juego recibe $Nx - r_1x_1 - \cdots - r_{n-1}x_{n-1} - (r_n - 1)x_n$ pesos. Si tomamos x de tal manera que esta última cantidad resulte ser igual a x_n, entonces se tendrá un juego justo en el cual Q recibe x_i pesos cuando ocurre el evento A_i, es decir se tendrá un juego justo que ofrece a Q las mismas posibilidades de ganar que el juego original. Por participar en este juego, Q debe pagar x pesos, en donde x es solución de la ecuación:

$$Nx - r_1x_1 - \cdots - r_{n-1}x_{n-1} - (r_n - 1)x_n = x_n$$

Resolviendo ésta, se obtiene:

$$x = \frac{r_1x_1+\cdots+r_nx_n}{r_1+\cdots+r_n} = \frac{r_1}{r_1+\cdots+r_n}x_1 + \cdots + \frac{r_n}{r_1+\cdots+r_n}x_n$$

$$= x_1P(A_1) + \cdots + x_nP(A_n)$$

Por lo tanto, el valor del juego para Q es $x = \sum_{i=1}^{n} x_iP(A_i)$.

9.1. Esperanza de variables aleatorias discretas

Consideremos un experimento aleatorio $\mathcal{E}$, con espacio muestral Ω, y cuyos posibles resultados sean equiprobables. Sea $X : \Omega \mapsto \mathbb{R}$ una variable aleatoria discreta cuyos posibles valores formen un conjunto finito $\{x_1, \ldots, x_n\}$. Entonces, los eventos $A_i = [X = x_i]$, con $1 \leq i \leq n$, constituyen una partición de Ω. Cada posible valor x_i de X puede pensarse como una cantidad que se recibe si ocurre el evento A_i. De manera que X representa un juego del tipo que se considera en el ejemplo 9.3. El valor de este juego, que puede entonces definirse como el **valor de X**, está dado por $\sum_{i=1}^{n} x_iP\left[X = x_i\right]$.

En general, dada una variable aleatoria discreta X, parece difícil encontrar "su valor" mediante la definición de un juego justo que ofrezca las mismas posibilidades de ganancia que el juego definido por X como se describió en el párrafo anterior. Si embargo, la fórmula que se obtiene en el caso particular considerado en el ejemplo 9.3 puede servir como motivación para la siguiente definición:

DEFINICIÓN 9.4 (**Esperanza de variables aleatorias discretas**). *Sea X una variable aleatoria discreta y sean $x_1, x_2, \ldots$ sus posibles valores. Se dice que X tiene esperanza finita si la serie $\sum_k |x_k| P[X = x_k]$ es convergente y, en ese caso, se define la esperanza de X, $E[X]$, mediante la relación:*

$$E[X] = \sum_k x_k P[X = x_k]$$

Los siguientes dos ejemplos ilustran dos de los métodos que se utilizan para evaluar la sumatoria $\sum_k x_k P[X = x_k]$ en el caso de variables aleatorias que admiten como posibles valores únicamente enteros no negativos.

EJEMPLO 9.5. *Sea X una variable aleatoria con distribución binomial de parámetros n y p. Encuentre $E[X]$.*

Solución

$E[X] = \sum_{x=0}^{n} x P[X = x] = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$

$= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x}$

$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} = np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x}$

$= np \sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{n-1-x} = np$

EJEMPLO 9.6. *Sea X una variable aleatoria con distribución Poisson de parámetro λ. Encuentre $E[X]$.*

Solución

$E[X] = \sum_{k=0}^{\infty} k P[X = k] = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{n} \frac{k \lambda^{k-1}}{k!}$

$= \lambda e^{-\lambda} \frac{d}{d\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \frac{d}{d\lambda} \left(e^{\lambda} - 1\right) = \lambda$

▲

La siguiente proposición establece un método, en ocasiones más simple que la aplicación directa de la definición, para calcular esperanzas de variables aleatorias discretas cuyos posibles valores son únicamente valores enteros.

Proposición 9.7. *Si X es una variable aleatoria discreta que toma únicamente valores enteros, entonces:*

$$\sum_{\{x:x>0\}} x f_X(x) = \sum_{x=1}^{\infty} P[X \geq x]$$

$$\sum_{\{x:x<0\}} |x| f_X(x) = \sum_{x=1}^{\infty} P[X \leq -x]$$

Demostración

$\sum_{\{x:x>0\}} x f_X(x) = \sum_{x=1}^{\infty} x P[X = x] = P[X = 1] + 2P[X = 2] + \cdots$

$= \sum_{x=1}^{\infty} P[X = x] + \sum_{x=2}^{\infty} P[X = x] + \cdots$

$= P[X \geq 1] + P[X \geq 2] + \cdots = \sum_{x=1}^{\infty} P[X \geq x]$

$\sum_{\{x:x<0\}} |x| f_X(x) = \sum_{x=-1}^{-\infty} |x| P[X = x] = P[X = -1] + 2P[X = -2] + \cdots$

$= \sum_{x=1}^{\infty} P[X = -x] + \sum_{x=2}^{\infty} P[X = -x] + \cdots$

$= P[X \leq -1] + P[X \leq -2] + \cdots = \sum_{x=1}^{\infty} P[X \leq -x]$

■

Corolario 9.8. *Sea X una variable aleatoria discreta que toma únicamente valores enteros. Entonces X tiene esperanza finita si y sólo si las series $\sum_{x=1}^{\infty} P[X \geq x]$ y $\sum_{x=1}^{\infty} P[X \leq -x]$ convergen. Además, en ese caso, se tiene:*

$$E[X] = \sum_{x=1}^{\infty} P[X \geq x] - \sum_{x=1}^{\infty} P[X \leq -x]$$

Ejemplo 9.9. *Sea X una variable aleatoria con distribución geométrica de parámetro p. Encuentre $E[X]$.*

Solución

$E[X] = \sum_{x=1}^{\infty} P[X \geq x] = \sum_{x=1}^{\infty} (1-p)^x = \frac{1-p}{p}$

9.2. Esperanza de variables aleatorias absolutamente continuas

Definición 9.10 (**Esperanza de variables aleatorias absolutamente continuas**). *Sea X una variable aleatoria absolutamente continua con función de densidad f_X. Se dice que X tiene esperanza finita si la integral $\int_{-\infty}^{\infty} |x| f_X(x)dx$ es finita y, en ese caso, se define la esperanza de X, $E[X]$, mediante la relación:*

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

Ejemplo 9.11. *Sea X una variable aleatoria con distribución gama de parámetros α y λ. Encuentre $E[X]$.*

Solución

$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\lambda x}dx = \frac{1}{\lambda\Gamma(\alpha)} \int_0^\infty y^\alpha e^{-y}dy = \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} = \frac{\alpha}{\lambda}$

Ejemplo 9.12. *Sea X una variable aleatoria con distribución normal de parámetros μ y σ^2. Encuentre $E[X]$.*

Solución

$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2\sigma^2}(x-\mu)^2}dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (y+\mu) e^{-\frac{1}{2\sigma^2}y^2}dy$

$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{1}{2\sigma^2}y^2}dy + \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}y^2}dy = \mu$

9.3. Algunas ideas erróneas

Eventualmente surgen algunas ideas erróneas con relación al concepto de esperanza. Dos de ellas provienen de que a la esperanza de una variable aleatoria X se le llama también su valor esperado, término que sugiere dos cosas, primero, que la esperanza de X coincide con alguno de sus posibles valores, segundo, que la esperanza de X es, efectivamente, el valor que se espera obtener al realizar el experimento aleatorio.

La primera de esas ideas erróneas es muy simple de aclarar, basta, por ejemplo, con considerar una variable aleatoria X con distribución binomial de parámetros n y p, en donde np no es un número entero. En ese caso, la esperanza de X, np, no coincide con alguno de sus posibles valores.

La segunda de esas ideas erróneas está relacionada con la primera pues si la esperanza de una variable aleatoria no coincide con alguno de sus posibles valores, entonces no se puede esperar obtener ese valor al realizar el experimento aleatorio. Sin embargo, si se precisa de manera adecuada lo que se entiende por "el valor que se espera obtener" esta segunda idea errónea contiene algo de verdad. Para aclarar esto, consideremos una variable aleatoria discreta X cuyo conjunto de posibles valores sea finito, digamos $\{x_1, \ldots, x_m\}$. Supongamos ahora que el experimento aleatorio respecto al cual está definida X se repite n veces, obteniéndose X_k como valor de X en la realización número k. El promedio $\bar{X}$ de los valores que se obtienen para X está dado entonces por $\bar{X} = \frac{X_1+\cdots+X_n}{n}$.

Por otra parte, en cada una de las realizaciones del experimento, el valor de X es alguno de sus posibles valores. Para $k \in \{1, \ldots, m\}$, sea n_k el número de veces que se obtiene, como valor de X, el resultado x_k. Se tiene entonces:

$X_1 + \cdots + X_n = n_1 x_1 + \cdots + n_m x_m$

De manera que:

$\bar{X} = \frac{n_1 x_1 + \cdots + n_m x_m}{n} = x_1 \frac{n_1}{n} + \cdots + x_m \frac{n_m}{n}$

Ahora bien, el cociente $\frac{n_k}{n}$ es la frecuencia relativa con la que X toma el valor x_k, de manera que, si tomamos la interpretación frecuencial de la probabilidad, se esperaría que, cuando n es grande, dicho cociente se parezca a la probabilidad $P[X = x_k]$, obteniéndose así:

$$\bar{X} \approx x_1 P[X = x_1] + \cdots + x_m P[X = x_m] = E[X]$$

De esta manera llegamos a la conclusión de que **la esperanza de X representa el promedio de los valores que toma la variable aleatoria cuando el experimento aleatorio se repite muchas veces**. La esperanza de X no es entonces el valor que se espera obtener para X, pero sí el valor que se espera obtener en promedio. Esta interpretación de la esperanza de una variable aleatoria X será analizada en el segundo volumen de este libro a la luz de la ley débil de los grandes números.

Obsérvese que, independientemente de que se tome el punto de vista frecuentista, la esperanza de una variable aleatoria discreta X, de acuerdo a su definición, es un promedio ponderado de sus diferentes posibles valores, de tal manera que al valor x_k se le asigna el peso $P[X = x_k]$.

Otra idea errónea que suele tenerse con relación al concepto de esperanza es que su valor es igual, o por lo menos está cercano, a un valor x de X para el cual $f_X(x)$ es un máximo. Esto ocurre con algunas de las distribuciones más conocidas, por ejemplo en los casos de las distribuciones binomial, Poisson y normal. Sin embargo, considérese por ejemplo el caso de una variable aleatoria X con distribución geométrica de parámetro $p = \frac{1}{100}$. El valor máximo de $f_X(x)$ se obtiene cuando $x = 0$, pero $E[X] = 99$. De la misma manera, considérese el caso de una variable aleatoria X con distribución exponencial de parámetro $\lambda = \frac{1}{100}$. El valor máximo de $f_X(x)$ se obtiene cuando $x = 0$, pero $E[X] = 100$.

El siguiente ejemplo muestra que incluso puede darse el caso en que la esperanza de una variable aleatoria coincida con su valor menos probable.

EJEMPLO 9.13. *Sea X una variable aleatoria con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{1}{N(3N+2)}(2N - x) & \text{si } x \in \{0, \ldots, N-1\} \\ \frac{1}{N(3N+2)}x & \text{si } x \in \{N, \ldots, 2N\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde N es un número entero positivo.

Obsérvese que el valor x de X para el cual la probabilidad $P[X = x]$ es mínima corresponde a $x = N$. Además, como se muestra en la figura de abajo, la gráfica de f_X es simétrica alrededor de $x = N$.

La simetría de la función f_X alrededor de $x = N$ hace que $E[X] = N$. En efecto:

$$E[X] = \frac{1}{N(3N+2)} \sum_{x=0}^{N-1} x(2N-x) + \frac{1}{N(3N+2)} \sum_{x=N}^{2N} x^2$$

$$= \frac{1}{6}\frac{(4N+1)(N-1)}{3N+2} + \frac{1}{6}\frac{(N+1)(14N+1)}{3N+2} = N$$

▲

Por último, debe mencionarse que si una variable aleatoria X representa la ganancia en un cierto juego, debe tenerse cuidado en la interpretación de su esperanza como lo que se tiene que pagar por el juego para que resulte justo. El siguiente ejemplo es ilustrativo al respecto.

Ejemplo 9.14 (**Paradoja de San Petersburgo**). *Consideremos un juego consistente en la realización de lanzamientos consecutivos de una moneda, de tal manera que el juego se termina en el momento en que se obtenga por primera vez águila o, en caso de no obtenerse águila previamente, en el momento en que se hayan hecho N lanzamientos, en donde N es un número entero positivo fijo. Supongamos que usted participa en dicho juego, estableciéndose que, si se obtiene águila por primera ocasión en el k-ésimo lanzamiento, entonces recibirá 2^k pesos. Llamando X a su ganancia en el juego, se tiene:*

$$f_X(x) = \begin{cases} \frac{1}{2^k} & \text{si } x = 2^k,\ k \in \{1, \ldots, N\} \\ \frac{1}{2^N} & \text{si } x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$E[X] = \sum_{k=1}^{N} 2^k \frac{1}{2^k} = N$$

Interpretando a la esperanza como el precio que se tiene que pagar por participar en el juego de tal manera que resulte justo, usted tendría que pagar N pesos por participar en él.

Consideremos ahora un juego específico, en donde $N = 1,000,000$. Según lo establecido anteriormente, pagando un millón de pesos por participar en el juego, resultaría justo. Si se estableciera que usted paga únicamente $500,000$ pesos por participar en dicho juego, resulta entonces que éste le sería favorable. ¿Aceptaría usted participar en esas condiciones?

Obsérvese que para obtener una ganancia neta al participar en el juego, se requeriría que se obtuviera águila por primera vez después del lanzamiento número 18 y la probabilidad de que esto ocurra es igual a $\frac{1}{2^{19}} + \frac{1}{2^{20}} + \cdots + \frac{1}{2^N} < \frac{1}{2^{18}} = 3.8147 \times 10^{-6}$. La conclusión parece ser entonces que, ante la probabilidad tan pequeña de obtener una ganancia neta, es prácticamente seguro que se perderían los $500,000$ pesos que se pagan y entonces el juego parecería más bien desfavorable para usted. ¿Es errónea entonces la interpretación que estamos dando a la esperanza?

La aparente contradicción entre la interpretación que estamos dando a la esperanza y el razonamiento previo se resuelve si recurrimos a la interpretación de la esperanza de una variable aleatoria como el promedio de los valores que toma esa variable aleatoria cuando el correspondiente experimento aleatorio se repite muchas veces, pues de esa manera lo justo del juego puede interpretarse en el sentido de que, en promedio, se gana lo mismo que se pierde al participar en él. De manera que, efectivamente, el juego resulta ser justo siempre y cuando se juegue un número considerablemente grande de veces.

En el caso del juego específico planteado en el ejemplo, pagando $500,000$ pesos por participar en él, se tendría un juego que le es favorable, siempre y cuando estuviera usted en posibilidad de jugarlo el número de veces suficiente que le permitiera, por lo menos, recuperar sus pérdidas de los juegos previos. Por ejemplo, dado que la probabilidad de obtener una ganancia neta en un juego es aproximadamente igual a $\frac{1}{2^{18}}$, lo cual significa que en promedio se obtendrá una ganancia neta en uno de cada 2^{18} juegos, el número de juegos que tendría usted que jugar para esperar tener una ganancia, después de haber compensado sus pérdidas, debería de ser considerablemente superior a 2^{18}. Evidentemente entra aquí, por lo menos, la limitación de la fortuna que usted posee, la cual seguramente no le permite tener la posibilidad de jugar tal cantidad de juegos pagando $500,000$ pesos por participar en cada uno de ellos. Pero, si dispusiera de tan fabulosa fortuna, a la larga estaría usted ganando, en promedio, $500,000$ pesos en cada juego.

9.4. Definición general de la Esperanza

Las definiciones 9.4 y 9.10 permiten calcular la esperanza de una variable aleatoria discreta o de una variable aleatoria absolutamente continua, pero no todas las variables aleatorias son de ese tipo, un caso así lo vimos en el ejemplo 8.32. Se hace necesario entonces contar con una definición general de la esperanza que permita su cálculo en cualquier caso y también para poder estudiar sus propiedades generales.

Una definición general de la esperanza de una variable aleatoria se obtiene con base en una generalización de la proposición 9.7, así como del corolario 9.7, las cuales se obtienen a continuación:

PROPOSICIÓN 9.15. *Si X es una variable aleatoria discreta no negativa con función de densidad f_X, entonces:*

$$\int_0^\infty P\left[X > x\right] dx = \sum_{\{x \in V_X : x > 0\}} x f_X(x)$$

en donde V_X es el conjunto de posibles valores de X.

Demostración

$\int_0^\infty P\left[X > x\right] dx = \int_0^\infty \sum_{\{y \in V_X : y > x\}} f_X(y) dx$

$= \int_0^\infty \sum_{\{y \in V_X : y > 0\}} I_{[0,y)}(x) f_X(y) dx = \sum_{\{y \in V_X : y > 0\}} \int_0^\infty I_{[0,y)}(x) f_X(y) dx$

$= \sum_{\{y \in V_X : y > 0\}} y f_X(y)$

■

COROLARIO 9.16. *Sea X una variable aleatoria discreta no negativa, entonces X tiene esperanza finita si y sólo si la integral $\int_0^\infty P\left[X > x\right] dx$ converge. Además, en ese caso, se tiene $E\left[X\right] = \int_0^\infty P\left[X > x\right] dx$.*

PROPOSICIÓN 9.17. *Si X es una variable aleatoria absolutamente continua no negativa con función de densidad f_X, entonces:*

$$\int_0^\infty P\left[X > x\right] dx = \int_0^\infty x f_X(x) dx$$

Demostración

$\int_0^\infty P\left[X > x\right] dx = \int_0^\infty \int_x^\infty f_X(y) dy dx$

$= \int_0^\infty \int_0^y f_X(y) dx dy = \int_0^\infty y f_X(y) dy$

■

COROLARIO 9.18. *Sea X una variable aleatoria absolutamente continua no negativa, entonces X tiene esperanza finita si y sólo si la integral $\int_0^\infty P[X > x]\,dx$ converge. Además, en ese caso, se tiene: $E[X] = \int_0^\infty P[X > x]\,dx$.*

Obsérvese que, dada cualquier variable aleatoria no negativa, la función $x \mapsto P[X > x]$ está acotada y es monótona, por lo tanto, de acuerdo con el teorema 41 del Apéndice, es Riemann integrable en cualquier intervalo finito, de manera que la integral:

$$\int_0^\infty P[X > x]\,dx$$

siempre está definida (pudiendo ser igual a ∞).

Los corolarios 9.16 y 9.18 muestran que la esperanza de una variable aleatoria no negativa X, de esperanza finita, se puede calcular, tanto en el caso discreto como en el caso absolutamente continuo, mediante la fórmula $E[X] = \int_0^\infty P[X > x]\,dx$. Además, la integral $\int_0^\infty P[X > x]\,dx$ está bien definida no sólo en los casos discreto y absolutamente continuo. Esto sugiere que se pueda utilizar la misma fórmula para definir la esperanza de cualquier variable aleatoria no negativa, como se hace a continuación:

DEFINICIÓN 9.19 (**Esperanza de variables aleatorias no negativas**). *Si X es una variable aleatoria no negativa, se define la esperanza de X, $E[X]$, mediante la fórmula:*

$$E[X] = \int_0^\infty P[X > x]\,dx$$

DEFINICIÓN 9.20. *Se dice que una variable aleatoria X tiene esperanza finita si $E[|X|] < \infty$.*

PROPOSICIÓN 9.21. $E[|X|] = \int_0^\infty P[X > x]\,dx + \int_{-\infty}^0 P[X < x]\,dx$

Demostración

$E[|X|] = \int_0^\infty P[|X| > x]\,dx = \int_0^\infty (P[X > x] + P[X < -x])\,dx$

$= \int_0^\infty P[X > x] + \int_0^\infty P[X < -x]\,dx$

$= \int_0^\infty P[X > x]\,dx + \int_{-\infty}^0 P[X < x]\,dx$

■

COROLARIO 9.22. *Una variable aleatoria X tiene esperanza finita si y sólo si las integrales $\int_0^\infty P[X > x]\,dx$ y $\int_{-\infty}^0 P[X < x]\,dx$ convergen.*

PROPOSICIÓN 9.23. *$E[X^+] = \int_0^\infty P[X > x]\,dx$ y $E[X^-] = \int_{-\infty}^0 P[X < x]\,dx$.*

Demostración

$E[X^+] = \int_0^\infty P[X^+ > x]\,dx = \int_0^\infty P[X > x]\,dx$

$E[X^-] = \int_0^\infty P[X^- > x]\,dx = \int_0^\infty P[X < -x]\,dx = \int_{-\infty}^0 P[X < x]\,dx$

■

COROLARIO 9.24. *Una variable aleatoria X tiene esperanza finita si y sólo si X^+ y X^- tienen esperanza finita.*

DEFINICIÓN 9.25. *Si X es una variable aleatoria de esperanza finita, se define la esperanza de X, $E[X]$, de la siguiente manera:*

$$E[X] = E[X^+] - E[X^-] = \int_0^\infty P[X > x]\,dx - \int_{-\infty}^0 P[X < x]\,dx$$

Obsérvese que, por la proposición 6.15 y el lema 6.17, $P[X < x] = F_X(x)$ excepto a lo más en un conjunto numerable, así que, por el teorema 43 del Apéndice, $\int_{-\infty}^0 P[X < x]\,dx = \int_{-\infty}^0 F_X(x)dx$. Además, $P[X > x] = 1 - F_X(x)$, así que entonces se tiene el siguiente resultado:

PROPOSICIÓN 9.26. *Sea X una variable aleatoria, entonces X tiene esperanza finita si y sólo si $\int_0^\infty [1 - F_X(x)]\,dx < \infty$ y $\int_{-\infty}^0 F_X(x)dx < \infty$ y, en ese caso, se tiene:*

$$E[X] = \int_0^\infty [1 - F_X(x)]\,dx - \int_{-\infty}^0 F_X(x)dx$$

EJEMPLO 9.27. *Sea X una variable aleatoria con distribución exponencial de parámetro λ y sea $M > 0$, Encuentre $E[\text{mín}(X, M)]$.*

Solución

Sea $Z = \text{mín}(X, M)$, entonces, como vimos en el ejemplo 8.32, Z es una variable aleatoria que no es ni discreta ni continua y se tiene:

$$F_Z(z) = \begin{cases} 0 & si\ z < 0 \\ P[X \leq z] & si\ 0 \leq z < M \\ 1 & si\ z \geq M \end{cases} = \begin{cases} 0 & si\ z < 0 \\ 1 - e^{-\lambda z} & si\ 0 \leq z < M \\ 1 & si\ z \geq M \end{cases}$$

Así que:

$E[Z] = \int_0^\infty [1 - F_Z(z)]\,dz - \int_0^\infty F_Z(-z-)dy = \int_0^M e^{-\lambda z}dz = \frac{1}{\lambda}\left(1 - e^{-\lambda M}\right)$

EJEMPLO 9.28. *Sea X una variable aleatoria con función de distribución dada por:*

$$F_X(x) = \begin{cases} 0 & si\ x < 0 \\ \frac{1}{5}(x+1) & si\ 0 \leq x < 1 \\ \frac{1}{5}(x+2) & si\ 1 \leq x < 2 \\ 1 & si\ x \geq 2 \end{cases}$$

Encuentre las esperanzas de X y X^2.

Solución

$E(X) = \int_0^2 P(X > x)dx = \int_0^1 \left[1 - \frac{1}{5}(x+1)\right] dx + \int_1^2 \left[1 - \frac{1}{5}(x+2)\right] dx$

$= \frac{1}{5}\int_0^1 (4-x)dx + \frac{1}{5}\int_1^2 (3-x)dx = 1$

$E(X^2) = \int_0^4 P(X^2 > y)dy = \int_0^4 P(X > \sqrt{y})dy$

$= \int_0^1 \left[1 - \frac{1}{5}(\sqrt{y}+1)\right] dy + \int_1^4 \left[1 - \frac{1}{5}(\sqrt{y}+2)\right] dy$

$= \frac{1}{5}\int_0^1 (4-\sqrt{y})dy + \frac{1}{5}\int_1^4 (3-\sqrt{y})dy = \frac{23}{15}$

PROPOSICIÓN 9.29. *Sea X una variable aleatoria no negativa y consideremos la descomposición $F_X = \alpha F_X^d + (1-\alpha) F_X^c$, en donde $\alpha \in [0,1]$, F_X^d es una función de distribución discreta y F_X^c es una función de distribución continua. Sean Y y Z dos variables aleatorias con funciones de distribución F_X^d y F_X^c, respectivamente. Entonces:*

$$E(X) = \alpha E(Y) + (1-\alpha) E(Z)$$

Demostración

$P[X > x] = 1 - F_X(x) = \alpha + 1 - \alpha - \alpha F_X^d(x) - (1-\alpha) F_X^c(x)$

$= \alpha \left[1 - F_X^d(x)\right] + (1-\alpha)\left[1 - F_X^c(x)\right]$

$= \alpha P[Y > x] + (1-\alpha) P[Z > x]$

Por lo tanto:

$\int_0^\infty P[X > x]\, dx = \alpha \int_0^\infty P[Y > x]\, dx + (1-\alpha)\int_0^\infty P[Z > x]\, dx$

Así que:

$E(X) = \alpha E(Y) + (1-\alpha) E(Z)$

■

COROLARIO 9.30. *Sea X una variable aleatoria y consideremos la descomposición $F_X = \alpha F_X^d + (1-\alpha) F_X^c$, en donde $\alpha \in [0,1]$, F_X^d es una función de distribución* *discreta y F_X^c es una función de distribución continua. Sean Y y Z dos variables aleatorias con funciones de distribución F_X^d y F_X^c, respectivamente. Entonces X tiene esperanza finita si y sólo si Y y Z tienen esperanza finita y, en ese caso, se tiene $E(X) = \alpha E(Y) + (1-\alpha) E(Z)$.*

Demostración

$\int_0^\infty [1 - F_X(x)]\, dx = \alpha \int_0^\infty \left[1 - F_X^d(x)\right] dx + (1-\alpha) \int_0^\infty [1 - F_X^c(x)]\, dx$

Así que:

$\int_0^\infty [1 - F_X(x)]\, dx < \infty$ si y sólo si $\int_0^\infty \left[1 - F_X^d(x)\right] dx < \infty$ y $\int_0^\infty [1 - F_X^c(x)]\, dx < \infty$.

$\int_{-\infty}^0 F_X(x)dx = \alpha \int_{-\infty}^0 F_X^d(x)dx + (1-\alpha) \int_{-\infty}^0 F_X^c(x)dx$

Así que $\int_{-\infty}^0 F_X(x)dx < \infty$ si y sólo si $\int_{-\infty}^0 F_X^d(x)dx < \infty$ y $\int_{-\infty}^0 F_X^c(x)dx < \infty$.

■

Cuando se tiene $\int_0^\infty P[X > x]\, dx = \infty$ y $\int_{-\infty}^0 P[X < x]\, dx < \infty$, se define $E[X] = \infty$, mientras que cuando $\int_0^\infty P[X > x]\, dx < \infty$ y $\int_{-\infty}^0 P[X < x]\, dx = \infty$, se define $E[X] = -\infty$. Cuando ambas integrales sean divergentes, entonces la esperanza de X no está definida.

Ejemplo 9.31. *Sea X una variable aleatoria con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{6}{\pi^2 x^2} & \text{si } x \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre $E[X]$.

Solución

Como X es una variable aleatoria no negativa, se tiene:

$E[X] = E[X^+] = \sum_{x=1}^\infty x f_X(x) = \frac{6}{\pi^2} \sum_{x=1}^\infty \frac{1}{x} = \infty$

Obsérvese que las probabilidades $P[X = x]$ decrecen a medida que crece x y sin embargo $E[X] = \infty$.

Ejemplo 9.32. *Consideremos una caminata aleatoria simple, tal como fue descrita en la sección 7.7, es decir una partícula se mueve a saltos sobre la recta real de tal manera que, comenzando en el origen, en cada intervalo de tiempo h, la partícula salta una distancia d hacia la derecha con probabilidad $\frac{1}{2}$ o una distancia d hacia la izquierda, también con probabilidad $\frac{1}{2}$.*

Sea R el paso en que ocurre el primer retorno al origen y Z el paso en el cual se alcanza un valor positivo por primera vez, es decir, R es el más pequeño número natural n tal que $S_n = 0$ y Z es el más pequeño número natural n tal que $S_n = d$, en donde S_n es la posición de la partícula en el tiempo nh. De acuerdo con los resultados de las secciones 7.7.3 y 7.7.5, se tiene:

$$P[R = 2m] = \begin{cases} \frac{1}{2m-1}\frac{1}{2^{2m}}\binom{2m}{m} & \textit{si } m \in \mathbb{N} \\ 0 & \textit{en otro caso} \end{cases}$$

$$P[Z = 2m-1] = \begin{cases} \frac{1}{2m-1}\frac{1}{2^{2m}}\binom{2m}{m} & \textit{si } m \in \mathbb{N} \\ 0 & \textit{en otro caso} \end{cases}$$

Ahora bien, con base en el corolario 6.48, se tiene:

$$\text{lím}_{m\rightsquigarrow\infty} \frac{\sqrt{\pi m}}{2^{2m}}\binom{2m}{m} = \text{lím}_{m\rightsquigarrow\infty} \frac{(2m)!\sqrt{\pi m}}{2^{2m}(m!)^2} = 1$$

Así que, existe $N \in \mathbb{N}$ tal que, para cualquier $m \geq N$, se tiene $\frac{\sqrt{\pi m}}{2^{2m}}\binom{2m}{m} > \frac{1}{2}$.

Por lo tanto:

$$\sum_{m=1}^{\infty} \frac{1}{2^{2m}}\binom{2m}{m} \geq \sum_{m=N}^{\infty} \frac{1}{2^{2m}}\binom{2m}{m}$$

$$= \sum_{m=N}^{\infty} \frac{\sqrt{\pi m}}{2^{2m}}\binom{2m}{m}\frac{1}{\sqrt{\pi m}} > \sum_{m=N}^{\infty} \frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{m}} > \frac{1}{2\sqrt{\pi}}\sum_{m=N}^{\infty} \frac{1}{m} = \infty$$

Así que:

$$E[R] = \sum_{m=1}^{\infty} \frac{2m}{2m-1}\frac{1}{2^{2m}}\binom{2m}{m} > \sum_{m=1}^{\infty} \frac{1}{2^{2m}}\binom{2m}{m} = \infty$$

$$E[Z] = \sum_{m=1}^{\infty} \frac{1}{2^{2m}}\binom{2m}{m} = \infty$$

Obsérvese que, definiendo, para $m \in \mathbb{N}$, $t_m = P[R = 2m] = P[Z = 2m-1]$, se tiene:

$$\frac{t_m}{t_{m+1}} = \frac{\frac{1}{2m-1}\frac{1}{2^{2m}}\binom{2m}{m}}{\frac{1}{2m+1}\frac{1}{2^{2m+2}}\binom{2m+2}{m+1}} = \frac{(2m+1)2^2}{2m-1}\frac{(2m)!(m+1)!(m+1)!}{(2m+2)!m!m!} = \frac{2m+2}{2m-1} > 1$$

Así que, la sucesión (t_m) es decreciente. Es decir, también en este caso las probabilidades $P[R = 2m]$ y $P[Z = 2m-1]$ decrecen a medida que crece m y sin embargo $E[R] = E[Z] = \infty$.

▲

Los resultados del último ejemplo complementan los resultados sobre caminatas aleatorias que obtuvimos en la sección 7.7. Ahí demostramos que, tanto la probabilidad de que haya una infinidad de retornos al origen, como la probabilidad de que haya una infinidad de pasos por $z = d$, es igual a 1. Ahora podemos agregar que, sin embargo, el número de pasos que se requieren, en promedio, para obtener el primer retorno al origen, o el primer paso por $z = d$, es ∞.

Ejemplo 9.33. *Sea X una variable con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{1}{|x|} & \text{si } x = \pm 2^k \text{ con } k \in \{2, 3, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

¿Está definida la esperanza de X?

Solución

$E[X^+] = E[X^-] = \sum_{k=2}^{\infty} 2^k \frac{1}{2^k} = \infty$. Por lo tanto, la esperanza de X no está definida.

Ejemplo 9.34. *Sea X una variable con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{6}{\pi^2 x^2} & \text{si } x \in \{1, 3, \ldots\} \cup \{-2, -4, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

¿Está definida la esperanza de X?

Solución

$E[X^+] = \sum_{k=0}^{\infty} (2k+1) \frac{6}{\pi^2 (2k+1)^2} = \frac{6}{\pi^2} \sum_{k=0}^{\infty} \frac{1}{2k+1} = \infty$

$E[X^-] = \sum_{k=1}^{\infty} 2k \frac{6}{\pi^2 (2k)^2} = \frac{3}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k} = \infty$

Por lo tanto, la esperanza de X no está definida. Esto a pesar de que la serie $\sum_{k=1}^{\infty} x_k f_X(x_k) = \frac{6}{\pi^2} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k}$ es convergente, en donde $x_k = (-1)^{k+1} k$.

9.5. Esperanza de funciones de variables aleatorias

Proposición 9.35. *Sea X una variable aleatoria discreta con función de densidad f_X y $g : \mathbb{R} \mapsto \mathbb{R}$ cualquier función boreliana no negativa. Entonces:*

$$\int_0^{\infty} P[g(X) > y]\, dy = \sum_{\{x \in V_X\}} g(x) f_X(x)$$

en donde V_X es el conjunto de posibles valores de X.

Demostración

a. $\int_0^{\infty} P[g(X) > y]\, dy = \int_0^{\infty} \sum_{\{x \in V_X : g(x) > y\}} f_X(x) dy$

$= \int_0^{\infty} \sum_{\{x \in V_X\}} I_{[0, g(x))}(y) f_X(x) dy = \sum_{\{x \in V_X\}} \int_0^{\infty} I_{[0, g(x))}(y) f_X(x) dy$

$= \sum_{\{x \in V_X\}} g(x) f_X(x)$

■

COROLARIO 9.36. *Sea X una variable aleatoria discreta con función de densidad f_X y $g : \mathbb{R} \mapsto \mathbb{R}$ cualquier función boreliana. Entonces, $g(X)$ tiene esperanza finita si y sólo si $\sum_{x \in V_X} |g(x)| f_X(x) < \infty$ y, en ese caso, se tiene:*

$E[g(X)] = \sum_{x \in V_X} g(x) f_X(x)$
en donde V_X es el conjunto de posibles valores de X.

PROPOSICIÓN 9.37. *Sea X una variable aleatoria absolutamente continua con función de densidad f_X y $g : \mathbb{R} \mapsto \mathbb{R}$ cualquier función boreliana no negativa (ver sección 5.7 y la proposición 6.5). Entonces:*

$$\int_0^\infty P[g(X) > y]\, dy = \int_{-\infty}^\infty g(x) f_X(x) dx$$

Demostración

a. $\int_0^\infty P[g(X) > y]\, dy = \int_0^\infty \int_{\{x \in \mathbb{R}: g(x) > y\}} f_X(x) dx dy$

$= \int_0^\infty \int_{-\infty}^\infty I_{[0,g(x))}(y) f_X(x) dx dy$

$= \int_{-\infty}^\infty \int_0^\infty I_{[0,g(x))}(y) f_X(x) dy dx$

$= \int_{-\infty}^\infty g(x) f_X(x) dx$

■

COROLARIO 9.38. *Sea X una variable aleatoria absolutamente continua con función de densidad f_X y $g : \mathbb{R} \mapsto \mathbb{R}$ cualquier función boreliana. Entonces, $g(X)$ tiene esperanza finita si y sólo si $\int_{-\infty}^\infty |g(x)| f_X(x) dx < \infty$ y, en ese caso, se tiene:*

$$E[g(X)] = \int_{-\infty}^\infty g(x) f_X(x) dx$$

EJEMPLO 9.39. *Sea X una variable aleatoria con distribución exponencial de parámetro λ y sea $M > 0$. Utilice el corolario 9.38 para calcular $E[\text{mín}(X, M]$.*

Solución

$E[\text{mín}(X, M] = \int_0^\infty \lambda \,\text{mín}(x, M) e^{-\lambda x} dx = \int_0^M \lambda x e^{-\lambda x} dx + \int_M^\infty \lambda M e^{-\lambda x} dx$

$= \frac{1}{\lambda}\left(1 - e^{-\lambda M} - \lambda M e^{-\lambda M}\right) + M e^{-\lambda M} = \frac{1}{\lambda}\left(1 - e^{-\lambda M}\right)$

EJEMPLO 9.40. *Sea X una variable aleatoria con distribución Poisson de parámetro λ. Encuentre la esperanza de $(1 + X)^{-1}$.*

Solución

$E\left[(1+X)^{-1}\right] = \sum_{k=0}^{\infty}(1+k)^{-1}\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{(k+1)!} = \frac{1}{\lambda}\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^{k+1}}{(k+1)!}$

$= \frac{1}{\lambda}\sum_{k=1}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!} = \frac{1}{\lambda}\left(1-e^{-\lambda}\right)$

EJEMPLO 9.41. *La demanda de periódicos que tiene un vendedor tiene distribución binomial con parámetros $n = 300$ y $p = \frac{1}{3}$. Cada periódico lo compra en 3 pesos y lo vende en 4 pesos, pero no se le reembolsa el costo de los periódicos que compre y no venda. ¿Cuántos periódicos debe de comprar de tal manera que el valor esperado de su ganancia sea máximo?*

Solución

Sea m el número de periódicos que compra, X el número de periódicos que vende y $G(m)$ la ganancia que obtiene el vendedor. Entonces:

$G(m) = \begin{cases} X - 3(m-X) & si\ X < m \\ m & si\ X \geq m \end{cases} = \begin{cases} 4X - 3m & si\ X < m \\ m & si\ X \geq m \end{cases}$

Así que:

$E\left[G(m)\right] = \sum_{k=0}^{m-1}(4k-3m)P\left[X=k\right] + m\sum_{k=m}^{n}P\left[X=k\right]$

$= \sum_{k=0}^{m-1}(4k-4m)P\left[X=k\right] + m\sum_{k=0}^{m-1}P\left[X=k\right] + m\sum_{k=m}^{n}P\left[X=k\right]$

$= m + 4\sum_{k=0}^{m-1}(k-m)P\left[X=k\right]$

Por lo tanto:

$E\left[G(m+1)\right] - E\left[G(m)\right]$

$= m + 1 + 4\sum_{k=0}^{m}(k-m-1)P\left[X=k\right] - m - 4\sum_{k=0}^{m-1}(k-m)P\left[X=k\right]$

$= 1 - 4\sum_{k=0}^{m}P\left[X=k\right] = 1 - 4P\left[X \leq m\right]$

Así que la función $G(m)$ crece mientras $P\left[X \leq m\right] < \frac{1}{4}$, después de lo cual decrece. Es decir, la ganancia es máxima cuando m es el más pequeño número entero tal que $P\left[X \leq m\right] \geq \frac{1}{4}$.

Por el teorema de de Moivre Laplace, se tiene:

$P\left[X \leq m\right] \approx P\left[Z \leq \frac{m+0.5-100}{\sqrt{\frac{200}{3}}}\right]$

en donde Z es una variable aleatoria con distribución normal estándar.

De acuerdo con la tabla de la distribución normal estándar, de la página 403, y definiendo $\Phi(z) = P\left[0 \leq Z \leq z\right]$*, se tiene:*

$\Phi(0.67) = P\left[0 \leq Z \leq 0.67\right] = 0.2486$

$\Phi(0.68) = P\left[0 \leq Z \leq 0.68\right] = 0.2518$

Además, de acuerdo con lo expuesto en la sección 8.3, para $z \in (0.67, 0.68)$*:*

$\Phi(z) \approx \Phi(0.67) + \frac{\Phi(0.68)-\Phi(0.67)}{0.68-0.67}\left(z - 0.67\right)$

$= 0.2486 + \frac{0.2518-0.2486}{0.68-0.67}\left(z - 0.67\right) = 0.2486 + 0.32\left(z - 0.67\right)$

Así que $\Phi(z) = 0.25$ *cuando* $z = 0.674375$*.*

Por lo tanto:

$P\left[Z \leq \frac{m+0.5-100}{\sqrt{\frac{200}{3}}}\right] \geq \frac{1}{4}$ *cuando* $\frac{m+0.5-100}{\sqrt{\frac{200}{3}}} \geq -0.674375.$

es decir, $m \geq 99.5 - 0.674375\sqrt{\frac{200}{3}} = 93.9938$

Por lo tanto, el valor esperado de la ganancia es máximo cuando $m = 94$*.*

Aunque este valor se obtuvo aproximando la distribución binomial mediante una distribución normal estándar, el resultado obtenido es exacto. En efecto, se tiene:

$P\left[X \leq 93\right] = \sum_{k=0}^{93} \binom{300}{k}\left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{300-k} = 0.21374$

$P\left[X \leq 94\right] = \sum_{k=0}^{94} \binom{300}{k}\left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{300-k} = 0.2515$

Para $m = 94$*, la ganancia esperada resulta ser:*

$E\left[G(94)\right] = 94 - 4\sum_{k=0}^{93}(94-k)\binom{300}{k}\left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{300-k} = 89.6642$

▲

En general, para encontrar la esperanza de una función f de una variable aleatoria X se puede recurrir a la definición general de la esperanza, para lo cual es necesario encontrar la función de distribución de $f(X)$ o, equivalentemente, las probabilidades $P\left[f(X) > y\right]$ para cualquier $y \in \mathbb{R}$.

Ejemplo 9.42. *Sea* X *una variable aleatoria con función de distribución dada por:*

$$F_X(x) = \begin{cases} 0 & si\ x < \frac{1}{2} \\ \frac{1}{10}x & si\ \frac{1}{2} \leq x < 2 \\ \frac{1}{10}x^2 & si\ 2 \leq x < 3 \\ 1 & si\ x \geq 3 \end{cases}$$

Encuentre la esperanza de $\ln X$*.*

Solución

$E(\ln X) = \int_0^{\ln 3} P(\ln X > y)dy - \int_{-\ln 2}^{0} P(\ln X < y)dy$

$= \int_0^{\ln 3} P(X > e^y)dy - \int_{-\ln 2}^{0} P(X < e^y)dy$

$= \int_1^3 \frac{1}{x} P(X > x)dx - \int_{\frac{1}{2}}^{1} \frac{1}{x} P(X < x)dx$

$= \int_1^2 \frac{1}{x}\left(1 - \frac{1}{10}x\right)dx + \int_2^3 \frac{1}{x}\left(1 - \frac{1}{10}x^2\right)dx - \int_{\frac{1}{2}}^{1} \frac{1}{10}dx$

$= \int_1^3 \frac{1}{x}dx - \frac{1}{10}\int_{\frac{1}{2}}^{2} dx - \frac{1}{10}\int_2^3 xdx = \ln 3 - \frac{2}{5}$

▲

Obsérvese que una variable aleatoria X puede tener una esperanza infinita, pero existir alguna función f tal que $f(X)$ tenga esperanza finita. Por ejemplo, si X es una variable aleatoria de esperanza infinita, entonces e^{-X} tiene esperanza finita.

9.6. Propiedades de la Esperanza

LEMA 9.43. *Sea X una variable aleatoria y $(X_n)_{n\in\mathbb{N}}$ una sucesión de variables aleatorias tales que $X_n \leq X_{n+1}$ para cualquier $n \in \mathbb{N}$ y $\lim_{n\rightsquigarrow\infty} X_n(\omega) = X(\omega)$ para cualquier $\omega \in \Omega$. Entonces, si para cada $n \in \mathbb{N}$, F_{X_n} es la función de distribución de X_n y F_X es la función de distribución de X, la sucesión $(F_{X_n}(x))$ es monótona no creciente y $\lim_{n\rightsquigarrow\infty} F_{X_n}(x) = F_X(x)$ para cualquier $x \in \mathbb{R}$.*

Demostración

Como $X_n \leq X_{n+1}$, se tiene, para cualquier $x \in \mathbb{R}$, $F_{X_{n+1}}(x) = P[X_{n+1} \leq x] \leq P[X_n \leq x] = F_{X_n}(x)$, así que la sucesión $(F_{X_n}(x))$ es monótona no creciente.

Ahora bien, como $\lim_{n\rightsquigarrow\infty} X_n(\omega) = X(\omega)$ para cualquier $\omega \in \Omega$, se tiene:

$\bigcap_{n=1}^{\infty} [X_n \leq x] = [X \leq x]$

para cualquier $x \in \mathbb{R}$, de manera que:

$F_X(x) = P[X \leq x] = \lim_{n\rightsquigarrow\infty} P[X_n \leq x] = \lim_{n\rightsquigarrow\infty} F_{X_n}(x)$

■

COROLARIO 9.44 (**Teorema de la convergencia monótona**). *Sea X una variable aleatoria no negativa y $(X_n)_{n\in\mathbb{N}}$ una sucesión de variables aleatorias no negativas tales que $X_n \leq X_{n+1}$ para cualquier $n \in \mathbb{N}$ y $\lim_{n\rightsquigarrow\infty} X_n(\omega) = X(\omega)$ para cualquier $\omega \in \Omega$. Entonces:*

$$E[X] = \lim_{n\rightsquigarrow\infty} E[X_n]$$

Demostración

Por el lema 9.43, se tiene que, para cualquier $x \in \mathbb{R}$, la sucesión $(1 - F_{X_n}(x))$ es monótona no decreciente. De manera que, por el teorema 47 del Apéndice, se tiene:

$E[X] = \int_0^\infty [1 - F_X(x)]\, dx = \lim_{n\rightsquigarrow\infty} \int_0^\infty [1 - F_{X_n}(x)]\, dx = \lim_{n\rightsquigarrow\infty} E[X_n]$ ■

LEMA 9.45. *Sea X una variable aleatoria no negativa, entonces existe una sucesión $(X_n)_{n\in\mathbb{N}}$ de variables aleatorias discretas no negativas tales que $X_n \leq X_{n+1}$ para cualquier $n \in \mathbb{N}$ y $\lim_{n\rightsquigarrow\infty} X_n(\omega) = X(\omega)$ para cualquier $\omega \in \Omega$.*

Demostración

Para cada $n \in \mathbb{N}$, sea $X_n = \sum_{m=0}^{\infty} \frac{m}{2^n} I_{\left[\frac{m}{2^n} \leq X < \frac{m+1}{2^n}\right]}$. Evidentemente X_n es una variable aleatoria discreta no negativa.

Dados $n \in \mathbb{N}$ y $\omega \in \Omega$, sea m el único número entero no negativo tal que $\frac{m}{2^n} \leq X(\omega) < \frac{m+1}{2^n}$. Entonces, como $X_n(\omega) = \frac{m}{2^n}$, se tiene $X(\omega) - \frac{1}{2^n} < X_n(\omega) \leq X(\omega)$, así que $\lim_{n\rightsquigarrow\infty} X_n(\omega) = X(\omega)$.

Ahora bien, como $\frac{2m}{2^{n+1}} \leq X(\omega) < \frac{2m+2}{2^{n+1}}$, se tiene que $\frac{2m}{2^{n+1}} \leq X(\omega) < \frac{2m+1}{2^{n+1}}$ o bien $\frac{2m+1}{2^{n+1}} \leq X(\omega) < \frac{2m+2}{2^{n+1}}$. En el primer caso, se tiene $X_{n+1}(\omega) = \frac{2m}{2^{n+1}} = X_n(\omega)$ mientras que en el segundo, se tiene $X_{n+1}(\omega) = \frac{2m+1}{2^{n+1}} > \frac{2m}{2^{n+1}} = X_n(\omega)$. Así que, en cualquier caso, $X_n(\omega) \leq X_{n+1}(\omega)$.

■

PROPOSICIÓN 9.46. *Sean X y Y dos variables aleatorias y c cualquier constante. Entonces:*

(i) *Si $P[X = c] = 1$, entonces $E[X] = c$.*
(ii) *Si X es no negativa, entonces $E[X] = 0$ si y sólo si $P[X = 0] = 1$.*
(iii) *Si X tiene esperanza finita, entonces cX también tiene esperanza finita y $E[cX] = cE[X]$.*
(iv) *Si $P[0 \leq X \leq Y] = 1$ y Y tiene esperanza finita, entonces X tiene esperanza finita y $E[X] \leq E[Y]$.*

(v) *Si X y Y tienen esperanza finita, entonces $X+Y$ también tiene esperanza finita y $E[X+Y]=E[X]+E[Y]$.*
(vi) *Si X y Y son independientes y tienen esperanza finita, entonces XY también tiene esperanza finita y $E[XY]=E[X]E[Y]$.*
(vii) *Si X y Y tienen esperanza finita y $P[X\leq Y]=1$, entonces $E[X]\leq E[Y]$.*
(viii) *Si X tiene esperanza finita, entonces $|E[X]|\leq E[|X|]$.*

Demostración

i. Si $P[X=c]=1$, entonces X es una variable aleatoria discreta cuyo único posible valor es c, de manera que se tiene $E[X]=cP[X=c]=c$.

ii. Supongamos $E[X]=0$, entonces $\int_0^\infty [1-F_X(x)]\,dx=E[X]=0$.

Sea $x>0$ tal que F_X es continua en x. Si $1-F_X(x)>0$, entonces existe $\varepsilon>0$ tal que $1-F_X>\varepsilon$ en una vecindad de x; así que se tendría $\int_0^\infty [1-F_X(x)]\,dx>0$. Por lo tanto, $F_X(x)=1$. Pero F_X es continua excepto en un conjunto numerable; de manera que, dada cualquier $y>0$ existe, x tal que $0<x<y$ y F_X es continua en x. Por lo tanto, $F_X(y)\geq F_X(x)=1$. Finalmente, como X es no negativa y F_X es continua por la derecha, se tiene:

$$F_X(x)=\begin{cases} 0 & \text{si } x<0 \\ 1 & \text{si } x\geq 0 \end{cases}$$

Así que, $P[X=0]=1$.

La otra implicación es corolario de *i*.

iii. Se tiene $(-X)^+=X^-$ y $(-X)^-=X^+$, de manera que, como X tiene esperanza finita, $(-X)^+$ y $(-X)^-$ tienen esperanza finita. Por lo tanto, $-X$ tiene esperanza finita y

$$E[-X]=E[(-X)^+]-E[(-X)^-]=E[X^-]-E[X^+]=-E[X]$$

Ahora bien, si $c>0$, se tiene:

$$\int_0^\infty [1-F_{cX}(x)]\,dx=\int_0^\infty P[cX>x]\,dx=\int_0^\infty P\left[X>\tfrac{x}{c}\right]dx$$

$$=c\int_0^\infty P[X>y]\,dy=cE[X^+]<\infty$$

$$\int_0^\infty F_{cX}(-x)dx=\int_0^\infty P[cX\leq x]\,dx=\int_0^\infty P\left[X\leq -\tfrac{x}{c}\right]dx$$

$$=c\int_0^\infty P[X\leq -y]\,dy=cE[X^-]<\infty$$

De manera que cX tiene esperanza finita y

$E[cX] = cE[X^+] - cE[X^-] = cE[X]$

Si $c < 0$, entonces $cX = -(-cX)$ tiene esperanza finita y se tiene:

$E[cX] = -E[(-c)X] = -(-c)E[X] = cE[X]$

iv. Para cualquier $x > 0$, se tiene:

$F_Y(x) = P[Y \leq x] \leq P[X \leq x] = F_X(x)$

de manera que $1 - F_X(x) \leq 1 - F_Y(x)$. Por lo tanto:

$\int_0^\infty [1 - F_X(x)]\, dx \leq \int_0^\infty [1 - F_Y(x)]\, dx < \infty$

Es decir, X tiene esperanza finita y $E[X] \leq E[Y]$.

v. Supongamos primero que X y Y son variables aleatorias discretas no negativas.

Sea V_X el conjunto de posibles valores de X, entonces:

$E[X + Y] = \int_0^\infty P[X + Y > z]\, dz$

$= \int_0^\infty \sum_{x \in V_X} P[X + Y > z, X = x]\, dz$

$= \sum_{x \in V_X} \int_0^\infty P[Y > z - x, X = x]\, dz$

$= \sum_{x \in V_X} \int_{-x}^\infty P[Y > y, X = x]\, dy$

$= \sum_{x \in V_X} \int_{-x}^0 P[X = x]\, dy + \int_0^\infty \sum_{x \in V_X} P[Y > y, X = x]\, dy$

$= \sum_{x \in V_X} xP[X = x] + \int_0^\infty P[Y > y]\, dy$

$= E[X] + E[Y]$

Sean ahora X y Y dos variables aleatorias no negativas. De acuerdo con el lema 9.45, existen dos sucesiones $(X_n)_{n\in\mathbb{N}}$ y $(Y_n)_{n\in\mathbb{N}}$ de variables aleatorias discretas no negativas tales que $X_n \leq X_{n+1}$ y $Y_n \leq Y_{n+1}$ para cualquier $n \in \mathbb{N}$ y $\lim_{n\rightsquigarrow\infty} X_n(\omega) = X(\omega)$ y $\lim_{n\rightsquigarrow\infty} Y_n(\omega) = Y(\omega)$ para cualquier $\omega \in \Omega$, lo cual implica $X_n + Y_n \leq X_{n+1} + Y_{n+1}$ para cualquier $n \in \mathbb{N}$ y $\lim_{n\rightsquigarrow\infty} (X_n + Y_n)(\omega) = (X + Y)(\omega)$ para cualquier $\omega \in \Omega$. De manera que, por el corolario 9.44, se tiene:

$E[X + Y] = \lim_{n\rightsquigarrow\infty} E[X_n + Y_n]$

$= \lim_{n\rightsquigarrow\infty} E[X_n] + \lim_{n\rightsquigarrow\infty} E[Y_n] = E[X] + E[Y]$

Finalmente, si X y Y son dos variables aleatorias cualesquiera de esperanza finita, se tiene:

$$|X+Y| = |X^+ + Y^+ - X^- - Y^-| \leq X^+ + Y^+ + X^- + Y^-$$

Pero, siendo X^+, X^-, Y^+ y Y^- variables aleatorias no negativas de esperanza finita, su suma también lo es, de manera que, por iv, $X+Y$ tiene esperanza finita. Además:

$$(X+Y)^+ - (X+Y)^- = X+Y = (X^+ + Y^+) - (X^- + Y^-)$$

Así que:

$$(X+Y)^+ + (X^- + Y^-) = (X^+ + Y^+) + (X+Y)^-$$

Por lo tanto:

$$E\left[(X+Y)^+\right] + E\left[X^- + Y^-\right] = E\left[(X+Y)^-\right] + E\left[X^+ + Y^+\right]$$

de lo cual se sigue:

$$E\left[X+Y\right] = E\left[(X+Y)^+\right] - E\left[(X+Y)^-\right] = E\left[X^+ + Y^+\right] - E\left[X^- + Y^-\right]$$

$$= E\left[X^+\right] + E\left[Y^+\right] - E\left[X^-\right] - E\left[Y^-\right] = \left(E\left[X^+\right] - E\left[X^-\right]\right) + \left(E\left[Y^+\right] - E\left[Y^-\right]\right)$$

$$= E\left[X\right] + E\left[Y\right].$$

vi. Supongamos primero que X y Y son variables aleatorias discretas no negativas.

Sea V_X el conjunto de posibles valores de X, entonces:

$$E\left[XY\right] = \int_0^\infty P\left[XY > z\right] dz$$

$$= \int_0^\infty \sum_{x \in V_X} P\left[XY > z, X = x\right] dz$$

$$= \sum_{\{x \in V_X : x > 0\}} \int_0^\infty P\left[Y > \tfrac{z}{x}, X = x\right] dz$$

$$= \sum_{\{x \in V_X : x > 0\}} \int_0^\infty xP\left[Y > y, X = x\right] dy$$

$$= \sum_{\{x \in V_X : x > 0\}} \int_0^\infty xP\left[X = x\right] P\left[Y > y\right] dy$$

$$= \sum_{\{x \in V_X : x > 0\}} xP\left[X = x\right] \int_0^\infty P\left[Y > y\right] dy$$

$$= E\left[X\right] E\left[Y\right]$$

Sean ahora X y Y dos variables aleatorias independientes no negativas. De acuerdo con el lema 9.45, existen dos sucesiones $(X_n)_{n\in\mathbb{N}}$ y $(Y_n)_{n\in\mathbb{N}}$ de variables aleatorias discretas no negativas tales que $X_n \leq X_{n+1}$ y $Y_n \leq Y_{n+1}$ para cualquier $n \in \mathbb{N}$ y $\lim_{n \rightsquigarrow \infty} X_n(\omega) = X(\omega)$ y $\lim_{n \rightsquigarrow \infty} Y_n(\omega) = Y(\omega)$ para cualquier $\omega \in \Omega$, lo cual implica $X_n Y_n \leq X_{n+1} Y_{n+1}$ para cualquier $n \in \mathbb{N}$ y $\lim_{n \rightsquigarrow \infty} (X_n Y_n)(\omega) = (XY)(\omega)$ para cualquier $\omega \in \Omega$. De manera que, por el corolario 9.44, se tiene:

$E[XY] = \lim_{n \rightsquigarrow \infty} E[X_n Y_n]$

$= \lim_{n \rightsquigarrow \infty} E[X_n] \lim_{n \rightsquigarrow \infty} E[Y_n] = E[X] E[Y]$

Además, de acuerdo con la demostración del lema 9.45, si, para cada $n \in \mathbb{N}$, definimos la función $f_n : \mathbb{R} \mapsto \mathbb{R}$ mediante la fórmula $f_n(x) = \frac{1}{2^n}[[2^n x]]$, en donde $[[2^n x]]$ denota al mayor entero menor o igual a $2^n x$, se puede elegir $X_n = f_n(X)$ y $Y_n = f_n(Y)$, de manera que, por la proposición 6.40, X_n y Y_n son independientes.

Finalmente, si X y Y dos variables aleatorias independientes cualesquiera de esperanza finita, se tiene:

$|XY| = |(X^+ - X^-)(Y^+ - Y^-)| \leq (X^+ + X^-)(Y^+ + Y^-)$

Pero, siendo $X^+ + X^-$ y $Y^+ + Y^-$ variables aleatorias no negativas de esperanza finita, su producto también lo es, de manera que, por *iv*, XY tiene esperanza finita.

Además, por la proposición 6.40, $X^+ = \text{máx}(X, 0)$ y $Y^+ = \text{máx}(Y, 0)$ son independientes y, de la misma manera, lo son X^+ y Y^-, X^- y Y^+ y X^- y Y^-. Por lo tanto:

$E[XY] = E[(X^+ - X^-)(Y^+ - Y^-)] = E[X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+]$

$= E[X^+Y^+] + E[X^-Y^-] - E[X^+Y^-] - E[X^-Y^+]$

$= E[X^+] E[Y^+] + E[X^-] E[Y^-] - E[X^+] E[Y^-] - E[X^-] E[Y^+]$

$= (E[X^+] - E[X^-])(E[Y^+] - E[Y^-]) = E[X] E[Y]$.

vii. Supongamos que X es no negativa y tiene esperanza finita, entonces, por *iv*, $E[X] \geq 0$.

Si $P[X \leq Y] = 1$, entonces $Y - X$ es una variable aleatoria no negativa de esperanza finita, por lo tanto, $E[Y - X] \geq 0$, lo cual, utilizando v, implica el resultado.

viii. $|E[X]| = |E[X^+] - E[X^-]| \leq E[X^+] + E[X^-] = E[|X|]$

■

Un razonamiento de inducción permite demostrar los siguientes dos corolarios:

COROLARIO 9.47. *Sean $X_1, \dots, X_n$ n variables aleatorias de esperanza finita, entonces $\sum_{k=1}^{n} X_k$ también tiene esperanza finita y $E\left[\sum_{k=1}^{n} X_k\right] = \sum_{k=1}^{n} E\left[X_k\right]$.*

COROLARIO 9.48. *Sean $X_1, \dots, X_n$ n variables aleatorias independientes de esperanza finita, entonces $\prod_{k=1}^{n} X_k$ también tiene esperanza finita y $E\left[\prod_{k=1}^{n} X_k\right] = \prod_{k=1}^{n} E\left[X_k\right]$.*

EJEMPLO 9.49. *Sea X una variable aleatoria con distribución hipergeométrica de parámetros r, s y n. Recordemos que esta distribución se presenta al tomar una muestra sin reemplazo de tamaño $n \leq r + s$ de una población formada por dos tipos de elementos, I y II, de tal manera que hay r elementos de tipo I y s de tipo II y definiendo X como el número de elementos de tipo I que se obtienen en la muestra. Para cada $i \in \{1, \dots, n\}$, definamos la variable aleatoria X_i de la siguiente manera:*

$$X_i = \begin{cases} 1 & \text{si el } i\text{-ésimo elemento de la muestra es de tipo I} \\ 0 & \text{si el } i\text{-ésimo elemento de la muestra es de tipo II} \end{cases}$$

entonces $E\left[X_i\right] = \frac{r}{r+s}$ y, como $X = X_1 + \cdots + X_n$, se tiene $E\left[X\right] = \frac{nr}{r+s}$.

EJEMPLO 9.50. *Consideremos una urna en la cual hay N bolas numeradas del 1 al N. Un juego consiste en ir sacando bolas al azar, una a una y sin reemplazo, hasta tener todas fuera, recibiendo un peso cada vez que el lugar de la elección coincida con el número de la bola seleccionada. Sea X la cantidad que se recibe al terminar el juego. Para cada $i \in \{1, \dots, N\}$, definamos la variable aleatoria X_i de la siguiente manera:*

$$X_i = \begin{cases} 1 & \text{si en la } i\text{-ésima elección se obtiene la bola número } i \\ 0 & \text{en otro caso} \end{cases}$$

entonces $E\left[X_i\right] = \frac{1}{N}$ y, como $X = X_1 + \cdots + X_N$, se tiene $E\left[X\right] = 1$.

EJEMPLO 9.51 (**Problema del colector de cupones**). *Se van colocando al azar bolas, una a una, en cualquiera de n cajas hasta que ninguna caja se encuentre vacía. ¿Cuál es la esperanza del número de bolas que se utilizan?*

Solución

Sea X el número de bolas que se utilizan y, para $j \in \{1, \dots, n\}$, sea X_j el número de bolas que se utilizan a partir del momento en que hay exactamente $j-1$ cajas ocupadas hasta que alguna bola queda colocada en una caja distinta a cualquiera de esas $j-1$ cajas. Se tiene entonces $X_1 = 1$ y, para $j \in \{2, \dots, n\}$, $X_j - 1$ es una variable aleatoria con distribución geométrica de parámetro $p = \frac{n-j+1}{n}$, así que $E\left[X_j\right] = 1 + \frac{1-p}{p} = 1 + \frac{j-1}{n-j+1} = \frac{n}{n-j+1}$

Por lo tanto:

$$E\left[X\right] = 1 + E\left[X_2\right] + E\left[X_3\right] + \cdots + E\left[X_n\right]$$

$= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1} = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right)$

EJEMPLO 9.52. *Una urna contiene N pares de tarjetas, de tal manera que, para cada $i \in \{1, \ldots, N\}$, hay un par de ellas marcadas con el número i. Si se seleccionan al azar m tarjetas, Encuentre la esperanza de a) el número de parejas que quedan dentro la urna y b) el número de parejas que quedan fuera de la urna. c) Consideremos una población en la cual hay 1000 matrimonios y supongamos que todas las personas de la población tienen la misma probabilidad de morir en cualquier momento dado. Al morir 500 de esas personas, ¿cuál es la esperanza del número de personas que quedan viudas?*

Solución

a. Sea X el número de pares que quedan en la urna y, para cada $i \in \{1, \ldots, N\}$, sea X_i una variable aleatoria que toma el valor 1 si el par de tarjetas marcadas con el número i queda en la urna después de las extracciones, y 0 en otro caso. Se tiene entonces:

$$E[X_i] = P[X_i = 1] = \frac{\binom{2N-2}{m}}{\binom{2N}{m}} = \frac{(2N-m)(2N-m-1)}{2N(2N-1)}$$

Así que:

$$E[X] = E\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N E[X_i] = \frac{(2N-m)(2N-m-1)}{2(2N-1)}$$

b. Sea Y el número de parejas que quedan fuera de la urna y, para cada $i \in \{1, \ldots, N\}$, sea Y_i una variable aleatoria que toma el valor 1 si el par de tarjetas marcadas con el número i queda fuera de la urna, y 0 en otro caso. Se tiene entonces:

$$E[Y_i] = P[Y_i = 1] = \frac{\binom{2}{2}\binom{2N-2}{m-2}}{\binom{2N}{m}} = \frac{m(m-1)}{2N(2N-1)}$$

Así que:

$$E[Y] = E\left[\sum_{i=1}^N Y_i\right] = \sum_{i=1}^N E[Y_i] = \frac{m(m-1)}{2(2N-1)}$$

c. Definamos el número de rompimientos como el número de parejas tales que uno de sus elementos queda fuera de la urna y el otro dentro, y sea Z el número de rompimientos, se tiene entonces $Z = N - X - Y$, así que:

$$E[Z] = N - E[X] - E[Y] = N - \frac{(2N-m)(2N-m-1)}{2(2N-1)} - \frac{m(m-1)}{2(2N-1)} = \frac{2N-m}{2N-1}m$$

Obsérvese que, si N es grande, para valores pequeños de m, esta esperanza es aproximadamente igual a m. Esto significa que se esperaría que los m elementos seleccionados fueran de parejas distintas.

En el caso de la población de matrimonios, el número de personas que quedan viudas es igual al número de rompimientos, de manera que:

$E[Z] = \frac{2N-m}{2N-1}m = \frac{2000-500}{2000-1}(500) = 375.19$

Digamos que una pareja es afectada si por lo menos uno de sus elementos queda fuera de la urna y sea U el número de parejas afectadas. Se tiene entonces $U = N - X$, así que:

$E[U] = N - \frac{(2N-m)(2N-m-1)}{2(2N-1)} = \frac{4N-m-1}{4N-2}m$

Obsérvese que, si N es grande, para valores pequeños de m esta esperanza es aproximadamente igual a m.

U es también el número de números distintos seleccionados.

En el caso de la población de matrimonios, se tiene:

$E[U] = \frac{4N-m-1}{4N-2}m = \frac{4000-500-1}{4000-2}(500) = 437.59$

$E[X] = \frac{(2N-m)(2N-m-1)}{2(2N-1)} = \frac{(2000-500)(2000-500-1)}{2(2000-1)} = 562.41$

$E[Y] = \frac{m(m-1)}{2(2N-1)} = \frac{500(500-1)}{2(2000-1)} = 62.406$

Ejemplo 9.53. *Una caja contiene n componentes eléctricos, de los cuales r están defectuosos. Se checa cada uno de los componentes en un orden aleatorio hasta que se encuentren los r defectuosos. ¿Cuál es la esperanza del número de chequeos que se realizan?*

Solución

Sea X el número de chequeos que se realizan y, para $i \in \{1, \ldots, n-r\}$, sea X_i una variable aleatoria que toma el valor 1 si la i-ésima lámpara que no está defectuosa se selecciona antes de que se hayan seleccionado todas las que sí lo están, y 0 en otro caso. Entonces se tiene:

$E[X_i] = P[X_i = 1] = \frac{r}{r+1}$

Así que:

$E[X] = E\left[r + \sum_{i=1}^{n-r} X_i\right] = r + \sum_{i=1}^{n-r} E[X_i] = r + \frac{(n-r)r}{r+1} = \frac{r(n+1)}{r+1}$

Ejemplo 9.54. *Sean $X_1, \ldots, X_n$ n variables aleatorias idénticamente distribuidas de esperanza finita y tales que $P\left[\sum_{k=1}^{n} X_k = 0\right] = 0$, entonces:*

$1 = E\left[\sum_{k=1}^{n} \frac{X_k}{\sum_{k=1}^{n} X_k}\right] = \sum_{k=1}^{n} E\left[\frac{X_k}{\sum_{k=1}^{n} X_k}\right] = nE\left[\frac{X_k}{\sum_{k=1}^{n} X_k}\right]$

Así que, $E\left[\frac{X_k}{\sum_{k=1}^{n} X_k}\right] = \frac{1}{n}$.

PROPOSICIÓN 9.55. *Sean $n \in \mathbb{N}$, X una variable aleatoria tal que X^n tiene esperanza finita y p un polinomio de grado n, entonces:*

$$E\left[p(X)\right] = p(0) + \int_0^\infty p'(x)\left[1 - F_X(x)\right]dx - \int_{-\infty}^0 p'(x)F_X\left(x\right)dx$$

Demostración

Para cualquier $r \in \mathbb{N}$ y $x \in \mathbb{R}$, se tiene $|x|^{r-1} \leq 1 + |x|^r$, así que $E\left[|X|^{r-1}\right] \leq 1 + E\left[|X|^r\right]$. Por lo tanto, si X^r tiene esperanza finita, entonces también X^{r-1} tiene esperanza finita.

Por otra parte, para $k \in \{1, \ldots, n\}$ y $y \in \mathbb{R}$, se tiene:

$$P\left[X^k \leq y\right] = \begin{cases} P\left[-y^{\frac{1}{k}} \leq X \leq y^{\frac{1}{k}}\right] & \text{si } y \geq 0 \text{ y } k \text{ es par} \\ P\left[X \leq y^{\frac{1}{k}}\right] & \text{si } y \geq 0 \text{ y } k \text{ es impar} \\ 0 & \text{si } y < 0 \text{ y } k \text{ es par} \\ P\left[X \leq -(-y)^{\frac{1}{k}}\right] & \text{si } y < 0 \text{ y } k \text{ es impar} \end{cases}$$

Así que:

$$E\left[X^k\right] = \begin{cases} \int_0^\infty \left[1 - F_X(y^{\frac{1}{k}}) + F_X(-y^{\frac{1}{k}})\right]dy & \text{si } k \text{ es par} \\ \int_0^\infty \left[1 - F_X(y^{\frac{1}{k}})\right]dy - \int_{-\infty}^0 F_X\left(-(-y)^{\frac{1}{k}}\right)dy & \text{si } k \text{ es impar} \end{cases}$$

$$= \begin{cases} \int_0^\infty kx^{k-1}\left[1 - F_X(x) + F_X(-x)\right]dx & \text{si } k \text{ es par} \\ \int_0^\infty kx^{k-1}\left[1 - F_X(x)\right]dx - \int_{-\infty}^0 kx^{k-1}F_X\left(x\right)dx & \text{si } k \text{ es impar} \end{cases}$$

$$= \int_0^\infty kx^{k-1}\left[1 - F_X(x)\right]dx - \int_{-\infty}^0 kx^{k-1}F_X\left(x\right)dx$$

Por lo tanto, si $p(x) = \sum_{k=0}^{n} a_k x^k$, se tiene:

$$E\left[p\left(X\right)\right] = p(0) + \sum_{k=1}^{n} a_k E\left[X^k\right]$$

$$= p(0) + \int_0^\infty \sum_{k=1}^{n} ka_k x^{k-1}\left[1 - F_X(x)\right]dx - \int_{-\infty}^0 \sum_{k=1}^{n} ka_k x^{k-1}F_X\left(x\right)dx$$

$$= p(0) + \int_0^\infty p'(x)\left[1 - F_X(x)\right]dx - \int_{-\infty}^0 p'(x)F_X\left(x\right)dx$$

■

Ejemplo 9.56. *Sea X una variable aleatoria con función de distribución dada por:*

$$F_X(x) = \begin{cases} 0 & si\ x < 0 \\ \frac{1}{5}(x+1) & si\ 0 \leq x < 1 \\ \frac{1}{5}(x+2) & si\ 1 \leq x < 2 \\ 1 & si\ x \geq 2 \end{cases}$$

Encuentre la esperanza de X^n para cualquier $n \in \mathbb{N}$.

Solución

$E(X^n) = \int_0^2 nx^{n-1}\left[1 - F_X(x)\right] dx$

$= \int_0^1 nx^{n-1}\left[1 - \frac{1}{5}(x+1)\right] dx + \int_1^2 nx^{n-1}\left[1 - \frac{1}{5}(x+2)\right] dx$

$= \frac{n}{5}\int_0^1 (4-x)x^{n-1}dx + \frac{2n}{5}\int_1^2 (3-x)x^{n-1}dx$

$= \frac{n}{5}\left(\frac{4}{n} - \frac{1}{n+1} + \frac{3}{n}2^n - \frac{3}{n} + \frac{1}{n+1} - \frac{1}{n+1}2^{n+1}\right)$

$= \frac{n}{5}\left(\frac{1}{n} + \frac{3}{n}2^n - \frac{1}{n+1}2^{n+1}\right) = \frac{1}{5} + \frac{1}{5}\frac{n+3}{n+1}2^n$

Proposición 9.57. *Sea X una variable aleatoria y consideremos la descomposición $F_X = \alpha F_X^d + (1-\alpha) F_X^c$, en donde $\alpha \in [0,1]$, F_X^d es una función de distribución discreta y F_X^c una función de distribución continua. Sean Y y Z dos variables aleatorias con funciones de distribución F_X^d y F_X^c, respectivamente. Entonces:*

$$E\left[g(X)\right] = \alpha E\left[g(Y)\right] + (1-\alpha) E\left[g(Z)\right]$$

para cualquier función boreliana no negativa.

Demostración

Denotemos por $\mathcal{B}$ a la σ-álgebra de Borel en $\mathbb{R}$ y recordemos que $\mathcal{B}$ es la σ-álgebra de subconjuntos de $\mathbb{R}$ generada por la familia de intervalos de la forma $(-\infty, x]$, en donde $x \in \mathbb{R}$.

Por la proposición 6.8, las funciones $\mu_X : \mathcal{B} \mapsto \mathbb{R}$, $\mu_Y : \mathcal{B} \mapsto \mathbb{R}$ y $\mu_Z : \mathcal{B} \mapsto \mathbb{R}$ definidas por $\mu_X(B) = P\left[X \in B\right]$, $\mu_Y(B) = P\left[Y \in B\right]$ y $\mu_Z(B) = P\left[Z \in B\right]$, respectivamente, son medidas de probabilidad. Por lo tanto, también la función $\mu : \mathcal{B} \mapsto \mathbb{R}$ definida por $\mu(B) = \alpha\mu_Y(B) + (1-\alpha)\mu_Z(B)$ es una medida de probabilidad.

Además, como $F_X = \alpha F_X^d + (1-\alpha) F_X^c$, $\mu_X((-\infty, x]) = \mu((-\infty, x])$ para cualquier $x \in \mathbb{R}$. Por lo tanto, aplicando la proposición 5.39, $\mu_X(B) = \mu(B)$ para cualquier $B \in \mathcal{B}$.

Sea $g : \mathbb{R} \mapsto \mathbb{R}$ una función boreliana no negativa. Para cada $m \in \mathbb{N}$, sea:

$$\varphi_m(x) = \begin{cases} \sum_{k=0}^{m2^m} \frac{k-1}{2^m} I_{\left\{y\in\mathbb{R}:\frac{k-1}{2^m}\leq g(y)<\frac{k}{2^m}\right\}}(x) & \text{si } g(x) < m \\ 0 & \text{si } g(x) \geq m \end{cases}$$

Dados $m \in \mathbb{N}$ y $x \in \mathbb{R}^n$, sea k el único número entero no negativo tal que $\frac{k-1}{2^m} \leq g(x) < \frac{k}{2^m}$. Entonces, como $\varphi_m(x) = \frac{k-1}{2^m}$, se tiene $g(x) - \frac{1}{2^m} < \varphi_m(x) \leq g(x)$, así que $\text{lím}_{m\rightsquigarrow\infty} \varphi_m(x) = g(x)$.

Ahora bien, como $\frac{2(k-1)}{2^{m+1}} \leq g(x) < \frac{2k}{2^{m+1}}$, se tiene que $\frac{2k-2}{2^{m+1}} \leq g(x) < \frac{2k-1}{2^{m+1}}$ o bien $\frac{2k-1}{2^{m+1}} \leq g(x) < \frac{2k}{2^{m+1}}$. En el primer caso, se tiene $\varphi_{m+1}(x) = \frac{2k-2}{2^{m+1}} = \frac{k-1}{2^m} = \varphi_m(x)$ mientras que en el segundo, se tiene $\varphi_{m+1}(x) = \frac{2k-1}{2^{m+1}} > \frac{2k-2}{2^{m+1}} = \varphi_m(x)$. Así que, en cualquier caso, $\varphi_m(\omega) \leq \varphi_{m+1}(\omega)$.

Así que, φ_m es una sucesión monótona no decreciente de funciones no negativas tales que $\text{lím}_{m\rightsquigarrow\infty} \varphi_m(x) = g(x)$ para cualquier $x \in \mathbb{R}^n$.

Para cada $m \in \mathbb{N}$, se tiene:

$$E\left[\varphi_m(X)\right] = E\left[\sum_{k=0}^{m2^m} \frac{k-1}{2^m} I_{\left\{y\in\mathbb{R}:\frac{k-1}{2^m}\leq g(y)<\frac{k}{2^m},g(y)<m\right\}}(X)\right]$$

$$= \sum_{k=0}^{m2^m} \frac{k-1}{2^m} E\left[I_{\left\{y\in\mathbb{R}:\frac{k-1}{2^m}\leq g(y)<\frac{k}{2^m},g(y)<m\right\}}(X)\right]$$

$$= \sum_{k=0}^{m2^m} \frac{k-1}{2^m} P\left[X \in \left\{y \in \mathbb{R} : \frac{k-1}{2^m} \leq g(y) < \frac{k}{2^m}, g(y) < m\right\}\right]$$

$$= \sum_{k=0}^{m2^m} \frac{k-1}{2^m} \mu_Y\left(\left\{y \in \mathbb{R} : \frac{k-1}{2^m} \leq g(y) < \frac{k}{2^m}, g(y) < m\right\}\right)$$

$$= \alpha \sum_{k=0}^{m2^m} \frac{k-1}{2^m} \mu_Y\left(\left\{y \in \mathbb{R} : \frac{k-1}{2^m} \leq g(y) < \frac{k}{2^m}, g(y) < m\right\}\right)$$

$$+ (1-\alpha) \sum_{k=0}^{m2^m} \frac{k-1}{2^m} \mu_Z\left(\left\{y \in \mathbb{R} : \frac{k-1}{2^m} \leq g(y) < \frac{k}{2^m}, g(y) < m\right\}\right)$$

$$= \alpha \sum_{k=0}^{m2^m} \frac{k-1}{2^m} P\left[Y \in \left\{y \in \mathbb{R} : \frac{k-1}{2^m} \leq g(y) < \frac{k}{2^m}, g(y) < m\right\}\right]$$

$$+ (1-\alpha) \sum_{k=0}^{m2^m} \frac{k-1}{2^m} P\left[Z \in \left\{y \in \mathbb{R} : \frac{k-1}{2^m} \leq g(y) < \frac{k}{2^m}, g(y) < m\right\}\right]$$

$$= \alpha \sum_{k=0}^{m2^m} \frac{k-1}{2^m} E\left[I_{\left\{y\in\mathbb{R}:\frac{k-1}{2^m}\leq g(y)<\frac{k}{2^m},g(y)<m\right\}}(Y)\right]$$

$$+ (1-\alpha) \sum_{k=0}^{m2^m} \frac{k-1}{2^m} E\left[I_{\left\{y\in\mathbb{R}:\frac{k-1}{2^m}\leq g(y)<\frac{k}{2^m},g(y)<m\right\}}(Z)\right]$$

$$= \alpha E\left[\sum_{k=0}^{m2^m} \frac{k-1}{2^m} I_{\left\{y\in\mathbb{R}:\frac{k-1}{2^m}\leq g(y)<\frac{k}{2^m},g(y)<m\right\}}(Y)\right]$$

$+ (1-\alpha) E\left[\sum_{k=0}^{m2^m} \frac{k-1}{2^m} I_{\left\{y\in\mathbb{R}: \frac{k-1}{2^m} \leq g(y) < \frac{k}{2^m}, g(y)<m\right\}}(Z)\right]$

$= \alpha E[\varphi_m(Y)] + (1-\alpha) E[\varphi_m(Z)]$

Así que, por el teorema de la convergencia monótona:

$E[g(X)] = \alpha E[g(Y)] + (1-\alpha) E[g(Z)]$

■

Corolario 9.58. *Sea X una variable aleatoria y consideremos la descomposición $F_X = \alpha F_X^d + (1-\alpha) F_X^c$, en donde $\alpha \in [0,1]$, F_X^d es una función de distribución discreta y F_X^c una función de distribución continua. Sean Y y Z dos variables aleatorias con funciones de distribución F_X^d y F_X^c, respectivamente. Entonces:*

$E[g(X)] = \alpha E[g(Y)] + (1-\alpha)[g(Z)]$

para cualquier función boreliana tal que $E[|g(X)|] < \infty$.

Demostración

$|g(X)| = g^+(X) + g^-(X)$, así que $E[|g(X)|] < \infty$ si y sólo si $E[g^+(X)] < \infty$ y $E[g^-(X)] < \infty$.

Además, por la proposición anterior:

$E[g^+(X)] = \alpha[g^+(Y)] + (1-\alpha)[g^+(Z)]$

$E[g^-(X)] = \alpha[g^-(Y)] + (1-\alpha)[g^-(Z)]$

Así que si $E[|g(X)|] < \infty$, entonces también $E[|g(Y)|] < \infty$ y $E[|g(Z)|] < \infty$.

Finalmente,

$E[g(X)] = E[g^+(X)] - E[g^-(X)]$

$= \alpha[E[g^+(Y)]] + (1-\alpha) E[g^+(Y)] - \alpha[E[g^-(Y)]] - (1-\alpha) E[g^-(Y)]$

$= \alpha(E[g^+(Y)] - E[g^-(Y)]) + (1-\alpha)(E[g^+(Z)] - E[g^-(Z)])$

$= \alpha E[g(Y)] + (1-\alpha)[g(Z)]$

■

Utilizando la notación del último corolario, obsérvese que se tiene:

$P[Y = x] = \frac{1}{\alpha} P[X = x]$

Así que:

$\alpha E\left[g(Y)\right] = \sum_{\{x\in\mathbb{R}:P[X=x]>0\}} g(x)P\left[X=x\right]$

Además, si F_X tiene derivada continua por pedazos, entonces F_X^c es absolutamente continua y se tiene:

$$E\left[g(X)\right] = \sum_{\{x\in\mathbb{R}:P[X=x]>0\}} g(x)P\left[X=x\right] + \int_{-\infty}^{\infty} g(x)F_X'(x)dx$$

Ejemplo 9.59. *Sea X una variable aleatoria con función de distribución dada por:*

$$F_X(x) = \begin{cases} 0 & si\ x < -\frac{1}{4}\pi \\ \frac{1}{8}\left(2-x^2\right) & si\ -\frac{1}{4}\pi \leq x < 0 \\ \frac{1}{4}\left(2+x\right) & si\ 0 \leq x < \frac{1}{4}\pi \\ 1 & si\ x \geq \frac{1}{4}\pi \end{cases}$$

Encuentre la esperanza de $\cos X$.

Solución

Se tiene:

$$P\left[X=x\right] = \begin{cases} \frac{1}{4} - \frac{\pi^2}{128} & si\ x = -\frac{\pi}{4} \\ \frac{1}{4} & si\ x = 0 \\ \frac{1}{2} - \frac{\pi}{16} & si\ x = \frac{\pi}{4} \\ 0 & en\ otro\ caso \end{cases}$$

$$F_X'(x) = \begin{cases} -\frac{x}{4} & si\ -\frac{\pi}{4} \leq x < 0 \\ \frac{1}{4} & si\ 0 \leq x < \frac{\pi}{4} \\ 0 & en\ otro\ caso \end{cases}$$

Así que:

$E\left[\cos X\right] = \sum_{\{x\in\mathbb{R}:P[X=x]>0\}} \cos x P\left[X=x\right] + \int_{-\infty}^{\infty} \cos x F_X'(x)dx$

$= \left(\frac{1}{4} - \frac{\pi^2}{128}\right)\cos\left(-\frac{\pi}{4}\right) + \frac{1}{4}\cos 0 + \left(\frac{1}{2} - \frac{\pi}{16}\right)\cos\frac{\pi}{4} - \int_{-\frac{\pi}{4}}^{0} \frac{x}{4}\cos x dx + \int_0^{\frac{\pi}{4}} \frac{1}{4}\cos x dx$

$= \left(\frac{1}{4} - \frac{\pi^2}{128}\right)\frac{\sqrt{2}}{2} + \frac{1}{4} + \left(\frac{1}{2} - \frac{\pi}{16}\right)\frac{\sqrt{2}}{2} - \frac{1}{4}\int_{-\frac{\pi}{4}}^{0} x\cos x dx + \frac{1}{4}\int_0^{\frac{\pi}{4}} \cos x dx$

$= \left(\frac{1}{4} + \frac{3\sqrt{2}}{8} - \frac{\sqrt{2}}{256}\pi^2 - \frac{\sqrt{2}}{32}\pi\right) + \left(\frac{\sqrt{2}}{4} + \frac{\sqrt{2}}{32}\pi - \frac{1}{4}\right)$

$= \frac{5\sqrt{2}}{8} - \frac{\sqrt{2}}{256}\pi^2 = 0.829361$

9.7. Varianza y demás momentos

Como ya vimos, la esperanza de una variable aleatoria X es un número que es aproximadamente igual al promedio de los valores que toma X cuando el correspondiente experimento aleatorio se repite muchas veces. Sin embargo, los valores particulares que toma X pueden diferir en mucho de su esperanza. ¿Qué tanto se alejan de su esperanza los valores que toma X? Esta pregunta tiene sentido no únicamente desde el punto de vista teórico, también lo tiene desde un punto de vista práctico pues siendo $E[X]$ aproximadamente igual al promedio de los valores que toma X, en un problema específico, $E[X]$ podría tomarse como una aproximación a los valores de X, cosa que funcionaría entonces muy bien en promedio, pero ¿qué pasaría con la aproximación al tener cada valor individual de X?, o en otras palabras, ¿qué error estaríamos cometiendo al aproximar un valor particular de X con su esperanza? Esto nos lleva a considerar la diferencia $X - E[X]$ y a establecer una medida para esta diferencia, lo cual podemos hacer de diferentes maneras.

Una manera podría consistir en considerar el promedio de los valores que toma la variable aleatoria $Y = |X - E[X]|$ y tomar ese promedio como una medida del alejamiento de los valores de X de su esperanza. Sin embargo, existe un problema con ese método. Para verlo claramente, observemos que el tomar el valor de $E[|X - E[X]|]$ como la medida del alejamiento de X de su esperanza, nos lleva a preguntarnos si, en este sentido, es la esperanza de X la mejor aproximación a los valores de X. Es decir, el número real que mejor aproxime a los valores de X debería ser uno para el cual la función $f(x) = E[|X - x|]$ alcanzara un valor mínimo. ¿Será $x = E[X]$ este valor? La respuesta es negativa. Consideremos, por ejemplo, una variable aleatoria con distribución exponencial de parámetro λ, se tiene entonces:

$$f(x) = E[|X - x|] = \int_0^\infty |y - x| \lambda e^{-\lambda y} dy = \int_0^x (x - y)\lambda e^{-\lambda y} dy + \int_x^\infty (y - x)\lambda e^{-\lambda y} dy$$

$$= \tfrac{1}{\lambda}(e^{-\lambda x} + \lambda x - 1) + \tfrac{1}{\lambda} e^{-\lambda x} = \tfrac{1}{\lambda}(2e^{-\lambda x} + \lambda x - 1)$$

Derivando esta expresión con respecto a x e igualando a cero se obtiene $x = \frac{1}{\lambda}\ln 2$, punto en el cual la función f alcanza un mínimo pues su segunda derivada en ese punto es positiva. De esta manera, no es en $x = E[X] = \frac{1}{\lambda}$ en donde la función f alcanza su valor mínimo.

Un valor x para el cual la función $f(x) = E[|X - x|]$ alcanza un valor mínimo tiene una caracterización interesante, la cual puede obtenerse del siguiente resultado:

PROPOSICIÓN 9.60. *Sea X una variable aleatoria de esperanza finita, entonces:*

$E[|X - x|] = E[|X|] + 2\int_0^x F_X(z)dz - x$ *para cualquier* $x \in \mathbb{R}$.

Demostración

Fijemos $x \in \mathbb{R}$ y sea $Y = |X - x|$, entonces, para $y \geq 0$, se tiene:

$F_Y(y) = P[|X - x| \leq y] = P[-y \leq X - x \leq y] = P[x - y \leq X \leq x + y]$

$= F_X(x + y) - F_X(x - y) + P[X = x - y]$

Así que:

$G(x) = E[|X - x|] = \int_0^\infty [1 - F_Y(y)]\, dy$

$= \int_0^\infty [1 - F_X(x + y) + F_X(x - y) - P[X = x - y]]\, dy$

$= \int_0^\infty [1 - F_X(x + y) + F_X(x - y)]\, dy = \int_x^\infty [1 - F_X(z)]\, dz + \int_{-\infty}^x F_X(z)dz$

$= \int_0^\infty [1 - F_X(z)]\, dz - \int_0^x [1 - F_X(z)]\, dz + \int_{-\infty}^0 F_X(z)dz + \int_0^x F_X(z)dz$

$= E[|X|] + 2\int_0^x F_X(z)dz - x$

■

Si X es una variable aleatoria continua, de la proposición 9.60 se sigue que la función $f(x) = E[|X - x|]$ alcanza un valor mínimo si y sólo si $F_X(x) = \frac{1}{2}$. Un número real x con esta propiedad es llamado una mediana de X. De manera general, se tiene la siguiente definición:

Definición 9.61 (**Mediana**). *Sea X una variable aleatoria cualquiera. Se dice que un número real m es mediana de X si $P[X \leq m] \geq \frac{1}{2}$ y $P[X \geq m] \geq \frac{1}{2}$.*

El ejercicio 9.39 desarrolla este concepto y en el ejercicio 9.41 se demuestra que $m \in \mathbb{R}$ es mediana de X si y sólo si la función $f(x) = E[|X - x|]$ alcanza su valor mínimo en m.. Por el momento basta con mencionar que uno de los problemas que se presenta con este concepto es que no siempre hay una única mediana, lo cual se ilustra con el siguiente ejemplo:

Ejemplo 9.62. *Sea X una variable aleatoria con distribución geométrica de parámetro p. Encuentre el conjunto de medianas de X.*

Solución

$P[X \leq n] = \sum_{x=0}^n P[X = x] = \sum_{x=0}^n p(1-p)^x = 1 - (1-p)^{n+1} \geq \frac{1}{2}$ *si y sólo si* $n \geq -\frac{\ln 2}{\ln(1-p)} - 1$.

$P[X \geq n] = \sum_{x=n}^\infty P[X = x] = \sum_{x=n}^\infty p(1-p)^x = (1-p)^n \geq \frac{1}{2}$ *si y sólo si* $n \leq -\frac{\ln 2}{\ln(1-p)}$.

De manera que si $-\frac{\ln 2}{\ln(1-p)}$ *es un número natural entonces* $m \in \mathbb{R}$ *es mediana de* X *si y sólo si* $m \in \left[-\frac{\ln 2}{\ln(1-p)} - 1, -\frac{\ln 2}{\ln(1-p)}\right]$*. Esto ocurre únicamente cuando* $p = 1-\left(\frac{1}{2}\right)^{\frac{1}{k}}$ *para algún número natural* k*, en cuyo caso,* m *es mediana de* X *si y sólo si* $m \in [k-1, k]$*.*

En el caso en que $-\frac{\ln 2}{\ln(1-p)}$ *no sea un número natural, es decir que* p *no sea de la forma* $p = 1 - \left(\frac{1}{2}\right)^{\frac{1}{k}}$ *con* $k \in \mathbb{N}$*, entonces* X *tiene una única mediana* m*, igual al más pequeño número entero no negativo* n *tal que* $n \geq -\frac{\ln 2}{\ln(1-p)} - 1$*.*

Por ejemplo si $p = 1 - \left(\frac{1}{2}\right)^{\frac{1}{8}}$*, entonces* $-\frac{\ln 2}{\ln(1-p)} = 8$*, así* m *es mediana de* X *si y sólo si* $m \in [7, 8]$*. En este caso,* $E[X] = \frac{1-p}{p} = 11.049$*, pero un valor para el cual* $E[|X-x|]$ *sea un mínimo es cualquier número real en el intervalo* $[7, 8]$*.*

Por otra parte, en cualquier caso, si $E[|X-x|]$ *es la medida del alejamiento entre* x *y los valores de* X*, se tendría que tomar como mejor aproximación de los valores de* X *a cualquiera de sus medianas en lugar de su esperanza.*

Proposición 9.63. *Sea* X *una variable aleatoria de varianza finita y definamos la función* $H : \mathbb{R} \mapsto \mathbb{R}$ *mediante la relación* $H(x) = E\left[(X-x)^2\right]$*. Entonces* H *alcanza su valor mínimo en* $x = E[X]$*.*

Demostración

$$H(x) = E\left[(X-x)^2\right] = E\left[X^2\right] - 2xE[X] + x^2$$

$$H'(x) = 2x - 2E[X]$$

Así que H alcanza su valor mínimo en $x = E[X]$.

∎

De acuerdo con la discusión anterior, podemos tomar a $E[|X-x|]$ como la medida del alejamiento entre x y los valores de X, en cuyo caso se tendría que tomar como mejor aproximación de los valores de X a cualquiera de sus medianas, o bien podemos tomar a $E\left[(X-x)^2\right]$ como la medida del alejamiento entre x y los valores de X, en cuyo caso la mejor aproximación de los valores de X es su esperanza. Las dos opciones son perfectamente aceptables, pero siendo la segunda la que conserva a la esperanza como la mejor aproximación de los valores de X, parece ser la mejor. Además, es la segunda la que históricamente surgió de manera natural, sobre todo en la formulación general de los teoremas que generalizan resultados como el de de Moivre, que se expuso en la sección 8.3, los cuales son conocidos como los teoremas límite y que han jugado un papel central en el desarrollo de la Teoría de la Probabilidad.

Lo anterior motiva las siguientes definiciones:

DEFINICIÓN 9.64 (**Varianza**). *Sea X una variable aleatoria de esperanza finita. Se define la varianza de X, $Var(X)$, mediante la relación:*

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

*A la raíz cuadrada no negativa de la varianza se le llama la **desviación estándar** de X.*

La varianza de una variable aleatoria mide entonces el alejamiento de los valores de X de su esperanza. También se acostumbra decir que la varianza es una medida de la dispersión de los valores de la variable aleatoria.

DEFINICIÓN 9.65 (**Varianza finita**). *Diremos que una variable aleatoria X tiene varianza finita si se cumplen las siguientes dos condiciones:*

(i) *X tiene esperanza finita.*
(ii) *$(X - E[X])^2$ tiene esperanza finita.*

PROPOSICIÓN 9.66. *Una variable aleatoria X tiene varianza finita si y sólo si X^2 tiene esperanza finita.*

Demostración

Se tiene $X^2 = (X - E[X])^2 + 2XE[X] - (E[X])^2$, así que si X tiene varianza finita, entonces X^2 tiene esperanza finita.

Supongamos ahora que X^2 tiene esperanza finita.

Se tiene $|X| \leq 1 + X^2$ y $(X - E[X])^2 = X^2 - 2XE[X] - (E[X])^2$. De manera que tanto X como $(X - E[X])^2$ tienen esperanza finita. Es decir, X tiene varianza finita.

PROPOSICIÓN 9.67. *Sea X una variable aleatoria de esperanza finita, entonces:*

$$Var(X) = E[X^2] - (E[X])^2$$

Demostración

Si X no tiene varianza finita entonces X^2 no tiene esperanza finita, así que se cumple la igualdad. Si X tiene varianza finita, entonces X^2 tiene esperanza finita y se tiene:

$$Var(X) = E\left[(X - E(X))^2\right] = E\left[X^2 - 2XE(X) + (E[X])^2\right]$$

$$= E[X^2] - 2(E[X])^2 + (E[X])^2$$

$$= E[X^2] - (E[X])^2$$

■

Ejemplo 9.68. *Sea X una variable aleatoria con distribución binomial de parámetros n y p. Encuentre la varianza de X.*

Solución

$E\left[X^2\right]=E\left[X(X-1)+X\right]=E\left[X(X-1)\right]+E\left[X\right]$

$=\sum_{x=0}^{n}x(x-1)P\left[X=x\right]+np=\sum_{x=0}^{n}x(x-1)\binom{n}{x}p^x(1-p)^{n-x}+np$

$=\sum_{x=0}^{n}x(x-1)\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}+np=\sum_{x=2}^{n}\frac{n!}{(x-2)!(n-x)!}p^x(1-p)^{n-x}+np$

$=n(n-1)p^2\sum_{x=2}^{n}\frac{(n-2)!}{(x-2)!(n-x)!}p^{x-2}(1-p)^{n-x}+np$

$=n(n-1)p^2\sum_{x=2}^{n}\binom{n-2}{x-2}p^{x-2}(1-p)^{n-x}+np$

$=n(n-1)p^2\sum_{x=0}^{n-2}\binom{n-2}{x}p^{x}(1-p)^{n-2-x}+np=n(n-1)p^2+np=n^2p^2+np(1-p)$

Así que:

$Var(X)=E\left[X^2\right]-\left(E\left[X\right]\right)^2=n^2p^2+np(1-p)-n^2p^2=np(1-p)$

Ejemplo 9.69. *Sea X una variable aleatoria con distribución Poisson de parámetro λ. Encuentre la varianza de X.*

Solución

$E\left[X^2\right]=E\left[X(X-1)+X\right]=E\left[X(X-1)\right]+E\left[X\right]$

$=\sum_{x=0}^{\infty}x(x-1)P\left[X=x\right]+\lambda=\sum_{x=0}^{\infty}x(x-1)\frac{\lambda^xe^{-\lambda}}{x!}+\lambda$

$=\lambda^2e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!}+\lambda=\lambda^2e^{-\lambda}e^{\lambda}+\lambda=\lambda^2+\lambda$

Así que:

$Var(X)=E\left[X^2\right]-\left(E\left[X\right]\right)^2=\lambda^2+\lambda-\lambda^2=\lambda$

Ejemplo 9.70. *Sea X una variable aleatoria con distribución geométrica de parámetro p. Encuentre la varianza de X.*

Solución

$E\left[X^2\right]=E\left[X(X-1)+X\right]=E\left[X(X-1)\right]+E\left[X\right]$

$=\sum_{x=0}^{\infty}x(x-1)p(1-p)^x+\frac{1-p}{p}$

$=p(1-p)^2\sum_{x=2}^{\infty}x(x-1)(1-p)^{x-2}+\frac{1-p}{p}$

$= p(1-p)^2 \sum_{x=2}^{\infty} \frac{d^2}{dp^2}(1-p)^x + \frac{1-p}{p} = p(1-p)^2 \frac{d^2}{dp^2} \sum_{x=2}^{\infty}(1-p)^x + \frac{1-p}{p}$

$= p(1-p)^2 \frac{d^2}{dp^2} \frac{(1-p)^2}{p} + \frac{1-p}{p} = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p}$

Así que:

$Var(X) = E[X^2] - (E[X])^2 = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p} - \frac{(1-p)^2}{p^2} = \frac{(1-p)^2}{p^2} + \frac{1-p}{p} = \frac{1-p}{p^2}$

EJEMPLO 9.71. *Sea X una variable aleatoria con función de densidad dada por*

$$f(x) = \begin{cases} \frac{1}{k(k+1)2^k} & si\ x = 2^k,\ k \in \mathbb{N} \\ 1 - \sum_{k=1}^{\infty} \frac{1}{k(k+1)2^k} & si\ x = 0 \\ 0 & en\ otro\ caso \end{cases}$$

Encuentre la varianza de X.

Solución

$E[X] = \sum_{k=1}^{\infty} \frac{2^k}{k(k+1)2^k} = \sum_{k=1}^{\infty} \left[\frac{1}{k} - \frac{1}{k+1}\right] = 1$

$E[X^2] = \sum_{k=1}^{\infty} \frac{2^{2k}}{k(k+1)2^k} = \sum_{k=1}^{\infty} \frac{2^k}{k(k+1)} = \infty$

Así que, aunque X tiene esperanza finita, su varianza es infinita.

EJEMPLO 9.72. *Sea X una variable aleatoria con distribución gama de parámetros α y λ. Encuentre la varianza de X.*

Solución

$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha+1} e^{-\lambda x} dx = \frac{1}{\lambda^2 \Gamma(\alpha)} \int_0^{\infty} y^{\alpha+1} e^{-y} dy$

$= \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} = \frac{\alpha(\alpha+1)}{\lambda^2}$

Así que:

$Var(X) = E[X^2] - (E[X])^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$

EJEMPLO 9.73. *Sea X una variable aleatoria con distribución normal de parámetros μ y σ^2. Encuentre la varianza de X.*

Solución

$Var(X) = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$

$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2\sigma^2}y^2} dy = \frac{2}{\sigma\sqrt{2\pi}} \int_0^{\infty} y^2 e^{-\frac{1}{2\sigma^2}y^2} dy = \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{\infty} z^{\frac{1}{2}} e^{-z^2} dz = \frac{2\sigma^2}{\sqrt{\pi}} \Gamma(\frac{3}{2}) = \sigma^2$

PROPOSICIÓN 9.74. *Sean X y Y dos variables aleatorias de varianza finita. Entonces, XY tiene esperanza finita.*

Demostración

Para cualquier par de números enteros x y y, se tiene $|xy| \leq x^2 + y^2$. Así que, $|XY| \leq X^2 + Y^2$.

Por lo tanto, XY tiene esperanza finita.

■

COROLARIO 9.75. *Si X y Y son dos variables aleatorias de varianza finita y a y b son dos números reales cualesquiera, entonces $aX + bY$ tiene varianza finita.*

Demostración

$(aX + bY)^2 = a^2X^2 + 2abXY + b^2Y^2$, así que, por las proposiciones 9.74 y 9.66, $aX + bY$ tiene varianza finita.

■

DEFINICIÓN 9.76 (**Covarianza**). *Sean X y Y dos variables aleatorias de varianza finita. Se define la covarianza de X y Y, $Cov(X, Y)$, mediante la relación:*

$$Cov(X,Y) = E\left[(X - E\left[X\right])\left(Y - E\left[Y\right]\right)\right] = E\left[XY\right] - E\left[X\right] E\left[Y\right]$$

PROPOSICIÓN 9.77. *Sean X y Y dos variables aleatorias independientes de varianza finita, entonces $Cov(X, Y) = 0$.*

Demostración

El resultado es inmediato pues X y Y son independientes y tienen esperanza finita, así que $E[XY] = E[X]\,E[Y]$.

■

El siguiente ejemplo muestra que la covarianza entre dos variables aleatorias puede ser cero sin que éstas sean independientes.

EJEMPLO 9.78. *Sea X una variable aleatoria con función de densidad dada por:*

$$f(x) = \begin{cases} \frac{1}{4} & si\ x \in \{-1, 1\} \\ \frac{1}{2} & si\ x = 0 \\ 0 & en\ otro\ caso \end{cases}$$

y sea Z una variable aleatoria, independiente de X, con distribución uniforme en el conjunto $\{-1, 1\}$. Definamos la variable aleatoria Y de la siguiente manera:

$Y = \begin{cases} Z & si\ X = 0 \\ 0 & en\ otro\ caso \end{cases}$

Se tiene entonces:

$f_Y(y) = P[Y = y] = P[Y = y, X = 0] + P[Y = y, X \neq 0]$

$= P[Z = y, X = 0] + P[Y = y, X \neq 0] = P[Z = y] P[X = 0] + I_{\{0\}}(y)P[X \neq 0]$

$= \frac{1}{4}I_{\{-1,1\}}(y) + \frac{1}{2}I_{\{0\}}(y) = \begin{cases} \frac{1}{4} & si\ y \in \{-1, 1\} \\ \frac{1}{2} & si\ y = 0 \\ 0 & en\ otro\ caso \end{cases}$

$f_{X,Y}(x, y) = P[X = x, Y = y] = P[X = x, Y = y, X = 0] + P[X = x, Y = y, X \neq 0]$

$= I_{\{0\}}(x)P[Z = y, X = 0] + I_{\{0\}}(y)I_{\{-1,1\}}(x)P[X = x]$

$= I_{\{0\}}(x)P[Z = y] P[X = 0] + I_{\{0\}}(y)I_{\{-1,1\}}(x)P[X = x]$

$= \frac{1}{4}I_{\{0\}}(x)I_{\{-1,1\}}(y) + \frac{1}{4}I_{\{0\}}(y)I_{\{-1,1\}}(x)$

$= \begin{cases} \frac{1}{4} & si\ x = 0, y \in \{-1, 1\}\ \ ó\ y = 0, x \in \{-1, 1\} \\ 0 & en\ otro\ caso \end{cases}$

Así que, por ejemplo, $P[X = 1, Y = 1] \neq P[X = 1] P[Y = 1]$*, de manera que* X *y* Y *no son independientes. Por otro lado,* $E[X] = E[Y] = E[XY] = 0$*, de manera que* $Cov(X, Y) = 0$*.*

PROPOSICIÓN 9.79. *Sean* X *y* Y *dos variables aleatorias de varianza finita y* a *y* b *dos números reales cualesquiera. Entonces* $Cov(aX, bY) = abCov(X, Y)$*.*

Demostración

$Cov(aX, bY) = E[aXbY]-E[aX] E[bY] = ab (E[XY] - E[X] E[Y]) = abCov(X, Y)$

■

PROPOSICIÓN 9.80. *Sean* $X, X_1, \ldots, X_n$ $n+1$ *variables aleatorias de esperanza finita. Entonces:*

(i) $Var(X) = 0$ *si y sólo si existe una constante* c *tal que* $P[X = c] = 1$*.*
(ii) $Var(aX + b) = a^2Var(X)$ *para cualesquiera constantes* a *y* b*.*
(iii) *Si* $X_1, \ldots, X_n$ *tienen varianza finita, entonces* $\sum_{i=1}^n X_i$ *también tiene varianza finita y:*
$Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + 2\sum_{\{i,j\in\{1,\ldots,n\}:i<j\}} Cov(X_i, Y_j)$

Demostración

i. $Var(X) = 0$ si y sólo si $E\left[(X - E(X))^2\right] = 0$, lo cual, por la proposición 9.46, ocurre si y sólo si $P\left[(X - E(X))^2 = 0\right] = 1$, es decir, $P[X = E(X)] = 1$.

ii. $Var(aX + b) = E\left[(aX - aE[X])^2\right] = a^2 E\left[(X - E[X])^2\right] = a^2 Var(X)$

iii. Que $\sum_{i=1}^n X_i$ tiene varianza finita se sigue del corolario 9.75 y un razonamiento de inducción. Además:

$$Var(\sum_{i=1}^n X_i) = E\left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n E[X_i]\right)^2\right] = E\left[\left(\sum_{i=1}^n (X_i - E[X_i])\right)^2\right]$$

$$= \sum_{i=1}^n E\left[(X_i - E[X_i])^2\right] + 2\sum_{\{i,j\in\{1,\ldots,n\}:i<j\}} E\left[(X_i - E[X_i])(X_j - E[X_j])\right]$$

$$= \sum_{i=1}^n Var(X_i) + 2\sum_{\{i,j\in\{1,\ldots,n\}:i<j\}} Cov(X_i, Y_j)$$

■

Corolario 9.81. *Sean $X_1, \ldots, X_n$ n variables aleatorias independientes y de varianza finita, entonces $\sum_{i=1}^n X_i$ también tiene varianza finita y $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i)$.*

Ejemplo 9.82. *Sea X una variable aleatoria con distribución normal de parámetros μ y σ^2. Encuentre la esperanza y la varianza de X.*

Solución

La variable aleatoria $Y = \frac{X-\mu}{\sigma}$ tiene distribución normal estándar y $X = \sigma Y + \mu$, así que:

$E[X] = \sigma E[Y] + \mu = \mu$

$Var(X) = \sigma^2 Var(Y) = \sigma^2$

Ejemplo 9.83. *n bolas se van colocando, una por una y al azar, en cualquiera de n cajas. Sea X el número de cajas que quedan vacías. Encuentre la esperanza y la varianza de X.*

Solución

Para cada $i \in \{1, \ldots, n\}$, definamos la variable aleatoria X_i de la siguiente manera:

$$X_i = \begin{cases} 1 & \text{si la } i\text{-ésima caja queda vacía} \\ 0 & \text{en otro caso} \end{cases}$$

Entonces, para $i, j \in \{1, \ldots, n\}$, se tiene:

$E[X_i] = \frac{(n-1)^n}{n^n} = \left(1 - \frac{1}{n}\right)^n$

$E[X_i^2] = E[X_i] = \left(1 - \frac{1}{n}\right)^n$

$E[X_iX_j] = \frac{(n-2)^n}{n^n} = \left(1 - \frac{2}{n}\right)^n$ *para* $i \neq j$

Así que:

$Var(X_i) = E[X_i^2] - (E[X_i])^2 = \left(1 - \frac{1}{n}\right)^n - \left(1 - \frac{1}{n}\right)^{2n}$

$Cov(X_i, X_j) = E[X_iX_j] - E[X_i]E[X_j] = \left(1 - \frac{2}{n}\right)^n - \left(1 - \frac{1}{n}\right)^{2n}$

para $i \neq j$.

y, como $X = \sum_{i=1}^n E[X_i]$, *entonces:*

$E[X] = \sum_{i=1}^n E[X_i] = n\left(1 - \frac{1}{n}\right)^n$

$Var(X) = \sum_{i=1}^n Var(X_i) + 2\sum_{\{i,j\in\{1,\ldots,n\},\ i<j\}} Cov(X_i, X_j)$

$= n\left(1 - \frac{1}{n}\right)^n - n\left(1 - \frac{1}{n}\right)^{2n} + 2\binom{n}{2}\left[\left(1 - \frac{2}{n}\right)^n - \left(1 - \frac{1}{n}\right)^{2n}\right]$

$= n\left(1 - \frac{1}{n}\right)^n - n\left(1 - \frac{1}{n}\right)^{2n} + n(n-1)\left(1 - \frac{2}{n}\right)^n - n(n-1)\left(1 - \frac{1}{n}\right)^{2n}$

$= n\left(1 - \frac{1}{n}\right)^n + n(n-1)\left(1 - \frac{2}{n}\right)^n - n^2\left(1 - \frac{1}{n}\right)^{2n}$

$= n\left[\left(1 - \frac{1}{n}\right)^n - \left(1 - \frac{2}{n}\right)^n\right] + n^2\left[\left(1 - \frac{2}{n}\right)^n - \left(1 - \frac{1}{n}\right)^{2n}\right]$

Obsérvese que si n *es grande, entonces* $E[X] \approx ne^{-1} \approx 0.3679n$, *así que, en promedio, aproximadamente el* $36.79\,\%$ *de las cajas quedan vacías.*

En ocasiones este problema se plantea de la siguiente forma: dadas n *tarjetas numeradas del* 1 *al* n, *se seleccionan, al azar y con reemplazo,* n *de ellas; entonces, si* n *es grande, en promedio, aproximadamente el* $36.79\,\%$ *de las tarjetas no son seleccionadas.*

PROPOSICIÓN 9.84 (**Desigualdad de Cauchy-Schwarz**). *Sean* X *y* Y *dos variables aleatorias cualesquiera, entonces:*

$$E[|XY|] \leq \sqrt{E[X^2]}\sqrt{E[Y^2]}$$

Además, si X *y* Y *tienen varianza finita, entonces* $|E[XY]| = \sqrt{E[X^2]}\sqrt{E[Y^2]}$ *si y sólo si existen constantes* a *y* b *tales que por lo menos una de ellas es distinta de cero y* $P[aX + bY = 0] = 1$.

Demostración

Si $E[X^2] = \infty$ o $E[Y^2] = \infty$ la desigualdad es obvia.

Supongamos ahora que $E[X^2] < \infty$ y $E[Y^2] < \infty$, es decir, que tanto X como Y tienen varianza finita.

Sea $\alpha = (E[Y^2])^{\frac{1}{2}}$ y $\beta = (E[X^2])^{\frac{1}{2}}$.

Si $\alpha = 0$, se tiene $E[X^2] = 0$, de manera que:

$P[|XY| = 0] \geq P[X = 0] = P[X^2 = 0] = 1$

Por lo tanto, $E[|XY|] = 0$. Así que se cumple la desigualdad.

De la misma manera, si $\beta = 0$, entonces $E[|XY|] = 0$. Así que se cumple la desigualdad.

Supongamos ahora que $\alpha > 0$ y $\beta > 0$.

Sabemos que $\alpha|X| - \beta|Y|$ tiene varianza finita y se tiene:

$0 \leq E\left[(\alpha|X| - \beta|Y|)^2\right] = \alpha^2 E[X^2] + \beta^2 E[Y^2] - 2\alpha\beta E[|XY|] = 2\alpha^2\beta^2 - 2\alpha\beta E[|XY|]$

Así que, $\alpha\beta - E[|XY|] \geq 0$. Es decir, $E[|XY|] \leq \alpha\beta$.

Para la segunda parte, supongamos primero que X y Y tienen varianza finita y que $|E[XY]| = \sqrt{E[X^2]}\sqrt{E[Y^2]}$.

Definiendo, como antes, $\alpha = (E[Y^2])^{\frac{1}{2}}$ y $\beta = (E[X^2])^{\frac{1}{2}}$, se tiene:

Si $\alpha = 0$ y $\beta = 0$, entonces $P[X = 0] = P[Y = 0] = 1$. Por lo tanto $P[X = 0, Y = 0] = 1$. De manera que, tomando en consideración que $P[X = 0, Y = 0] \leq P[X + Y = 0]$, se tiene $P[X + Y = 0] = 1$. Es decir, se tiene el resultado deseado con $a = b = 1$.

Si $\alpha \neq 0$ ó $\beta \neq 0$ se tienen los siguientes dos casos:

Si $E[XY] > 0$, entonces:

$0 \leq E\left[(\alpha X - \beta Y)^2\right] = 2\alpha^2\beta^2 - 2\alpha\beta E[XY] = 0$

Así que, $E\left[(\alpha X - \beta Y)^2\right] = 0$, de lo cual se sigue $P[\alpha X - \beta Y = 0] = 1$

Es decir, se tiene el resultado deseado con $a = \alpha$ y $b = -\beta$.

Si $E[XY] < 0$, entonces:

$0 \leq E\left[(\alpha X + \beta Y)^2\right] = 2\alpha^2\beta^2 + 2\alpha\beta E[XY] = 0$

Así que, $E\left[(\alpha X + \beta Y)^2\right] = 0$, de lo cual se sigue $P[\alpha X + \beta Y = 0] = 1$.

Es decir, se tiene el resultado deseado con $a = \alpha$ y $b = \beta$.

Finalmente, supongamos que existen constantes a y b tales que por lo menos una de ellas es distinta de cero y $P[aX + bY = 0] = 1$. Supongamos, por ejemplo, que $a \neq 0$, entonces $P\left[X = -\frac{b}{a}Y\right] = 1$. Así que:

$(E[XY])^2 = \frac{b^2}{a^2}(E[Y^2])^2 = E\left[\left(-\frac{b}{a}Y\right)^2\right]E[Y^2] = E[X^2]E[Y^2]$

∎

Corolario 9.85. *Sean X y Y dos variables aleatorias de varianza finita. Entonces:*

$$|Cov(X,Y)| \leq \sqrt{Var(X)}\sqrt{Var(Y)}$$

Además, la igualdad se cumple si y sólo si existen constantes a, b y c tales que a y b no son ambas cero y $P[aX + bY = c] = 1$.

Demostración

Utilizando la proposición 9.84, se tiene:

$|Cov(X,Y)| = |E[(X - E[X])(Y - E[Y])]| \leq E[|X - E[X]|\,|Y - E[Y]|]$

$\leq \sqrt{E\left[(X - E[X])^2\right]}\sqrt{E\left[(Y - E[Y])^2\right]} = \sqrt{Var(X)}\sqrt{Var(Y)}$

Si la igualdad se cumple, entonces se tiene

$|E[(X - E[X])(Y - E[Y])]| = \sqrt{E\left[(X - E[X])^2\right]}\sqrt{E\left[(Y - E[Y])^2\right]}$.

De manera que, nuevamente por la proposición 9.84, existen constantes a y b tales que no son ambas cero y $P[a(X - E[X]) + b(Y - E[Y]) = 0] = 1$. Es decir:

$P[aX + bY = c] = 1$

en donde $c = aE[X] + bE[Y]$.

Supongamos ahora que existen constantes a, b y c tales que a y b no son ambas cero y

$P[aX + bY = c] = 1$. Entonces $E[aX + bY - c] = 0$, de lo cual se sigue $c = E[aX + bY]$. De manera que se tiene $P[a(X - E[X]) + b(Y - E[Y]) = 0] = 1$. Así que, por la proposición 9.84, se tiene:

$$|Cov(X,Y)| = |E[(X - E[X])(Y - E[Y])]| = \sqrt{Var(X)}\sqrt{Var(Y)}$$

■

DEFINICIÓN 9.86 (**Momento de orden n**). *Sea X una variable aleatoria y n un entero no negativo. Si $|X|^n$ tiene esperanza finita, se dice que X tiene momento de orden n finito y a la cantidad $E[X^n]$ se le llama el momento de orden n de X.*

EJEMPLO 9.87. *Para cualquier $n \in \mathbb{N}$, encuentre el momento de orden n de una variable aleatoria con distribución normal estándar.*

Solución

Sea X una variable aleatoria con distribución normal estándar. Para cualquier $n \in \mathbb{N}$, se tiene:

$$E[|X|^n] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} |x|^n e^{-\frac{1}{2}x^2}dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^n e^{-\frac{1}{2}x^2}dx$$

$$= \frac{2^{\frac{n}{2}}}{\sqrt{\pi}}\int_0^{\infty} y^{\frac{n-1}{2}}e^{-y}dy = \frac{2^{\frac{n}{2}}}{\sqrt{\pi}}\Gamma(\tfrac{n+1}{2}) < \infty$$

Así que X tiene momento de orden n finito.

$$E[X^n] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^n e^{-\frac{1}{2}x^2}dx = \begin{cases} 0 & si\ n\ es\ impar \\ \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^n e^{-\frac{1}{2}x^2}dx & si\ n\ es\ par \end{cases}$$

$$= \begin{cases} 0 & si\ n\ es\ impar \\ \frac{2^{\frac{n}{2}}}{\sqrt{\pi}}\int_0^{\infty} y^{\frac{n-1}{2}}e^{-y}dy & si\ n\ es\ par \end{cases} = \begin{cases} 0 & si\ n\ es\ impar \\ \frac{2^{\frac{n}{2}}}{\sqrt{\pi}}\Gamma(\frac{n+1}{2}) & si\ n\ es\ par \end{cases}$$

$$= \begin{cases} 0 & si\ n\ es\ impar \\ 1 \cdot 3 \cdot 5 \cdots (n-1) & si\ n\ es\ par \end{cases}$$

PROPOSICIÓN 9.88. *Si una variable aleatoria X tiene momento de orden n finito, entonces, para cualquier entero no negativo $m \leq n$, X tiene momento de orden m finito.*

Demostración

Para cualquier $r \in \mathbb{N}$ y $x \in \mathbb{R}$, se tiene $|x|^{r-1} \leq 1 + |x|^r$, así que $E\left[|X|^{r-1}\right] \leq 1 + E[|X|^r]$. Por lo tanto, si X tiene momento de orden r finito, entonces también tiene momento de orden $r-1$ finito, lo cual prueba el resultado.

■

EJEMPLO 9.89. *Sea X una variable aleatoria con función de densidad f dada por:*

$$f(x) = \begin{cases} \frac{c}{x^{r+2}} & si\ x \in \mathbb{N} \\ 0 & en\ otro\ caso \end{cases}$$

en donde r es un entero no negativo y $c = \frac{1}{\sum_{x=1}^{\infty} \frac{1}{x^{r+2}}}$.

Sea n un entero no negativo. Si $n \leq r$, entonces:

$E[|X|^n] = c\sum_{x=1}^{\infty} \frac{x^n}{x^{r+2}} = c\sum_{x=1}^{\infty} \frac{1}{x^{(r-n)+2}} \leq c\sum_{x=1}^{\infty} \frac{1}{x^2} < \infty$

Así que X tiene momento de orden n finito.

Si $n > r$, entonces:

$E[|X|^n] = c\sum_{x=1}^{\infty} \frac{x^n}{x^{r+2}} = c\sum_{x=1}^{\infty} \frac{x^{n-r}}{x^2} \geq c\sum_{x=1}^{\infty} \frac{1}{x} = \infty$

Así que X no tiene momento de orden n finito.

9.8. Desigualdad de Chebyshev

PROPOSICIÓN 9.90. *Sea X cualquier variable aleatoria y ε cualquier número real positivo, entonces:*

$$P[|X| \geq \varepsilon] \leq \tfrac{1}{\varepsilon} E[|X|]$$

Demostración

$E[|X|] = \int_0^\infty \left[1 - F_{|X|}(x)\right] dx = \int_0^\varepsilon \left[1 - F_{|X|}(x)\right] dx + \int_\varepsilon^\infty \left[1 - F_{|X|}(x)\right] dx$

$\geq \int_0^\varepsilon \left[1 - F_{|X|}(x)\right] dx = \int_0^\varepsilon P[|X| > x]\, dx \geq \int_0^\varepsilon P[|X| \geq \varepsilon]\, dx = \varepsilon P[|X| \geq \varepsilon]$

■

COROLARIO 9.91. *Sea X cualquier variable aleatoria y ε cualquier número real positivo, entonces:*

$$P[|X| \geq \varepsilon] \leq \tfrac{1}{\varepsilon^2} E[X^2]$$

Demostración

$P[|X| \geq \varepsilon] = P[X^2 \geq \varepsilon^2] \leq \frac{1}{\varepsilon^2} E[X^2]$

■

COROLARIO 9.92 (**Desigualdad de Chebyshev**). *Sea X cualquier variable aleatoria de esperanza finita y ε cualquier número real positivo, entonces:*

$$P[|X - E[X]| \geq \varepsilon] \leq \tfrac{1}{\varepsilon^2} Var[X]$$

EJEMPLO 9.93. *Sea X una variable aleatoria con distribución binomial de parámetros n y p.*

Tomando $n = 16$, $p = \frac{1}{2}$ *y* $\varepsilon = 2$, *la desigualdad de Chebyshev establece:*

$P\left[|X - 8| < 1\right] \geq 0$

de lo cual no se obtiene información. Además:

$P\left[|X - 8| < 2\right] = P\left[7 \leq X \leq 9\right] = \frac{1}{2^{16}} \sum_{k=7}^{9} \binom{16}{k} = \frac{35750}{65536} = 0.5455$

Es decir, la cota que establece la desigualdad de Chebyshev está muy lejos del valor exacto.

Para valores de n *grandes, el teorema de de Moivre Laplace establece:*

$P\left[|X_n - np| > x\sqrt{npq}\right] \approx \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}y^2} dy$

mientras que la desigualdad de Chebyshev nos dice:

$P\left[|X_n - np| > x\sqrt{npq}\right] \leq \frac{1}{x^2}$

De manera que si $x \leq 1$ *la desigualdad de Chebyshev no da información y la cota que establece puede ser bastante mala, por ejemplo:*

$P\left[|X_n - np| > 0.2\sqrt{npq}\right] \approx \frac{2}{\sqrt{2\pi}} \int_{0.2}^\infty e^{-\frac{1}{2}y^2} dy = 0.8415$

Cuando x *es muy grande, la desigualdad de Chebyshev establece una cota cercana a cero, la cual está cercana al valor exacto de la probabilidad, sin embargo, aún en ese caso el error relativo que se comete, tomando la cota de la desigualdad de Chebyshev por la verdadera probabilidad, es muy grande ya que:*

$\lim_{x \rightsquigarrow \infty} \frac{\frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}y^2} dy}{\frac{1}{x^2}} = \lim_{x \rightsquigarrow \infty} \frac{\frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}}{\frac{2}{x^3}} = \frac{1}{\sqrt{2\pi}} \lim_{x \rightsquigarrow \infty} \frac{x^3}{e^{\frac{1}{2}x^2}}$

$= \frac{1}{\sqrt{2\pi}} \lim_{x \rightsquigarrow \infty} \frac{3x}{e^{\frac{1}{2}x^2}} = \frac{3}{\sqrt{2\pi}} \lim_{x \rightsquigarrow \infty} \frac{1}{xe^{\frac{1}{2}x^2}} = 0$

Es decir, el valor aproximado que da la desigualdad de Chebyshev es bastante más grande que el valor exacto.

Como conclusión podemos decir que, para el caso de una distribución binomial, con parámetro n *grande, es preferible recurrir al teorema de de Moivre Laplace para aproximarla.*

Ejemplo 9.94. *Sea* X *una variable aleatoria con distribución geométrica de parámetro* $p = \frac{1}{10}$. *Para* $\varepsilon = 10$ *la desigualdad de Chebyshev establece* $P\left[|X - 8| < 2\right] \geq 0.1$, *pero:*

$P\left[|X - 9| < 10\right] = P\left[X \leq 18\right] = \frac{1}{10} \sum_{k=0}^{18} \left(\frac{9}{10}\right)^k = 0.8649$

Es decir, la cota que establece la desigualdad de Chebyshev está muy lejos del valor exacto.

EJEMPLO 9.95. *Sea X una variable aleatoria con distribución exponencial de parámetro $\lambda = 1$. Para $\varepsilon = 1$ la desigualdad de Chebyshev establece $P[|X-1|<1] \geq 0$, pero:*

$P[|X-1|<1] = P[X<2] = \int_0^2 e^{-x}dx = 0.8647$

Es decir, la cota que establece la desigualdad de Chebyshev está muy lejos del valor exacto.

▲

A pesar de que en general la cota que establece la desigualdad de Chebyshev es mala, ésta puede alcanzarse, como se muestra en el siguiente ejemplo:

EJEMPLO 9.96. *Sea X una variable aleatoria que admite como posibles valores -1, 0 y 1 con probabilidades p, $1-2p$ y p, respectivamente, en donde $0<p<1$. Se tiene entonces, $E[X]=0$ y $Var(X)=2p$, así que, la desigualdad de Chebyshev establece $P[|X| \geq \varepsilon] \leq \frac{2p}{\varepsilon^2}$ para cualquier $\varepsilon > 0$. Para valores de ε mayores que 1 pero cercanos a 1 y valores de p cercanos a $\frac{1}{2}$, la cota que establece la desigualdad de Chebyshev es cercana a 1. Por ejemplo, para $\varepsilon = 1.01$ y $p = 0.49$, se tiene $\frac{2p}{\varepsilon^2} = 0.9607$, pero $P[|X| \geq \varepsilon] = 0$ para cualquier $\varepsilon > 1$. De manera que en este caso la cota que establece la desigualdad de Chebyshev es bastante burda. Sin embargo, para $\varepsilon = 1$, se tiene $P[|X| \geq \varepsilon] = 2p$, es decir, la cota se alcanza.*

▲

La desigualdad de Chebyshev puede utilizarse para resolver algunos problemas de estimación. Sin embargo, como en muchos casos la cota que establece es muy burda, el método en general no es muy bueno, tal como se muestra en los siguientes ejemplos:

EJEMPLO 9.97. *Supongamos que para decidir si una moneda está balanceada, la lanzamos $10,000$ veces y rechazamos la moneda como balanceada si el número de soles que se obtienen difiere de $5,000$ en k o más, en donde k es un número por determinar, de tal manera que, estando balanceada, la probabilidad de rechazar la moneda sea a lo más 0.01. Si llamamos X al número de soles en los $10,000$ lanzamientos y suponemos que la moneda está balanceada, la desigualdad de Chebyshev establece:*

$P[|X-5,000| \geq k] \leq \frac{2,500}{k^2}$

De manera que si tomamos k de tal forma que $\frac{2,500}{k^2} \leq 0.01$, es decir $k \geq 500$, la probabilidad de rechazar una moneda balanceada es menor que 0.01.

El procedimiento con cualquier $k \geq 500$, digamos $k = 500$, es bueno en tanto que satisface la condición de rechazar una moneda balanceada con probabilidad menor o igual a 0.01*. Sin embargo, veamos el resultado que se obtiene utilizando el teorema de de Moivre Laplace, de acuerdo al cual se tiene:*

$$P\left[|X - 5,000| < k\right] = P\left[-\frac{k}{50} < \frac{X-5000}{50} < \frac{k}{50}\right] \approx \frac{1}{\sqrt{2\pi}} \int_{-\frac{k-0.5}{50}}^{\frac{k-0.5}{50}} e^{-\frac{1}{2}x^2} dx \geq 0.99$$

De manera que tomando $\frac{k-0.5}{50} \geq 2.575$, es decir $k \geq 130$, se tiene:

$$P\left[|X - 5,000| < k\right] \geq 0.99$$

Así que basta con tomar $k \geq 130$ para que la probabilidad de rechazar una moneda balanceada sea menor que 0.01*.*

Obviamente, tomando $k = 500$ se satisface el requerimiento de que la probabilidad de rechazar una moneda balanceada sea menor que 0.01*, sin embargo, de esta manera el margen dentro del cual no se rechaza la moneda es más amplio, así que la probabilidad de aceptar una moneda no balanceada resulta bastante más alta que tomando $k = 130$. En efecto, si p es la probabilidad real de que la moneda caiga sol, tomando $k = 500$, la probabilidad $f(p)$ de que la moneda se acepte está dada por:*

$$f(p) = P\left[4500 < X < 5500\right] = P\left[\frac{45-100p}{\sqrt{p(1-p)}} < \frac{X-np}{\sqrt{npq}} < \frac{55-100p}{\sqrt{p(1-p)}}\right]$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{\frac{45.005-100p}{\sqrt{p(1-p)}}}^{\frac{54.995-100p}{\sqrt{p(1-p)}}} e^{-\frac{1}{2}x^2} dx$$

Mientras que, tomando $k = 130$, la probabilidad $g(p)$ de que la moneda se acepte está dada por:

$$g(p) = P\left[4870 < X < 5130\right] = P\left[\frac{48.7-100p}{\sqrt{p(1-p)}} < \frac{X-np}{\sqrt{npq}} < \frac{51.3-100p}{\sqrt{p(1-p)}}\right]$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{\frac{48.705-100p}{\sqrt{p(1-p)}}}^{\frac{51.295-100p}{\sqrt{p(1-p)}}} e^{-\frac{1}{2}x^2} dx$$

La gráfica siguiente muestra f en línea sólida y g punteada, observándose la notable diferencia entre ellas para algunos valores de p. Por ejemplo, $f(0.46) = 0.977055$, mientras que $g(0.46) = 2.85912 \times 10^{-8}$.

Ejemplo 9.98. *Una población muy grande está compuesta de dos clases de peces. Para estimar la proporción de peces de cada una de las clases, tomamos una muestra y la proporción de peces de tipo I que se obtenga en la muestra la tomamos como la proporción de peces de tipo I en la población. ¿De qué tamaño debe de ser la muestra para garantizar que la probabilidad de cometer un error menor que* 0.01 *con esta estimación sea mayor o igual que* 0.95*?*

Solución

Si p *es la proporción de peces de tipo I en la población,* n *el tamaño de la muestra y* X *el número de peces de tipo I que se obtienen en la muestra, entonces* X *tiene aproximadamente distribución binomial y se están buscando los valores de* n *para los cuales* $P\left[\left|\frac{X}{n}-p\right|<0.01\right]\geq 0.95$.

La desigualdad de Chebyshev establece:

$$P\left[\left|\frac{X}{n}-p\right|\geq 0.01\right]\leq \frac{np(1-p)}{(0.01n)^2}\leq \frac{2500}{n}$$

De manera que tomando $\frac{2500}{n}\leq 0.05$*, es decir* $n\geq 50000$*, se obtiene el resultado deseado.*

Por otra parte, el teorema de de Moivre Laplace establece:

$$P\left[\left|\frac{X}{n}-p\right|<0.01\right]=P\left[-\frac{0.01n}{\sqrt{npq}}\leq \frac{X-np}{\sqrt{npq}}<\frac{0.01n}{\sqrt{npq}}\right]$$

$$\approx \frac{1}{\sqrt{2\pi}}\int_{-\frac{.01n}{\sqrt{npq}}}^{\frac{.01n}{\sqrt{npq}}} e^{-\frac{1}{2}x^2}dx \geq \frac{1}{\sqrt{2\pi}}\int_{-\frac{.01n}{\frac{1}{2}\sqrt{n}}}^{\frac{.01n}{\frac{1}{2}\sqrt{n}}} e^{-\frac{1}{2}x^2}dx$$

De manera que tomando $\frac{0.01n}{\frac{1}{2}\sqrt{n}} \geq 1.96$*, es decir* $n \geq 9604$*, se obtiene el resultado deseado, lo cual muestra que la desigualdad de Chebyshev da un resultado muy burdo.*

▲

Por lo expuesto anteriormente, pareciera que, al establecer estimaciones muy burdas, la desigualdad de Chebyshev no tiene gran utilidad. Sin embargo, esto no es así pues ésta permite la demostración de resultados fundamentales en la Teoría de la Probabilidad. Pafnuty Lvovich Chebyshev la demostró en 1867 en un artículo en el cual mostró que el teorema de Bernoulli (ver proposición 7.4) se puede generalizar para obtener lo que se conoce actualmente como la Ley Débil de los Grandes Números, la cual fue uno de los resultados que hicieron avanzar fuertemente las investigaciones alrededor de los llamados **teoremas límite** de la Teoría de la Probabilidad.

Recordemos que el teorema de Bernoulli establece que si p es la probabilidad de ocurrencia de un evento A, en un determinado experimento aleatorio, entonces, llamando X_n al número de veces que ocurre A al realizar n repeticiones independientes del experimento, se tiene:

$$\lim_{n\rightsquigarrow\infty} P\left[\left|\tfrac{X_n}{n} - p\right| > \varepsilon\right] = 0$$

para cualquier número $\varepsilon > 0$.

Utilizando la desigualdad de Chebyshev, este resultado se demuestra como sigue:

$P\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right] = P\left[|X_n - np| > n\varepsilon\right] = P\left[|X_n - E\left[X_n\right]| > n\varepsilon\right]$

$\leq P\left[|X_n - E\left[X_n\right]| \geq n\varepsilon\right] \leq \frac{1}{n^2\varepsilon^2} Var\left[X_n\right] = \frac{np(1-p)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}$

Así que:

$\lim_{n\rightsquigarrow\infty} P\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right] \leq \lim_{n\rightsquigarrow\infty} \frac{p(1-p)}{n\varepsilon^2} = 0$

En el año 1800, Siméon Denis Poisson demostró una generalización del teorema de Bernoulli al considerar una sucesión indefinida de ensayos independientes de tal forma que la probabilidad de éxito en el ensayo n es igual a p_n. Entonces, llamando X_n al número de éxitos que se obtienen en los primeros n ensayos, se tiene:

$$\lim_{n\rightsquigarrow\infty} P\left[\left|\tfrac{X_n}{n} - \tfrac{1}{n}\sum_{j=1}^{n} p_j\right| > \varepsilon\right] = 0$$

para cualquier número $\varepsilon > 0$.

Utilizando la desigualdad de Chebyshev, este resultado se demuestra como sigue:

$P\left[\left|\frac{X_n}{n}-\frac{1}{n}\sum_{j=1}^{n}p_j\right|>\varepsilon\right]=P\left[\left|X_n-\sum_{j=1}^{n}p_j\right|>n\varepsilon\right]=P\left[\left|X_n-E\left[X_n\right]\right|>n\varepsilon\right]$

$\leq P\left[\left|X_n-E\left[X_n\right]\right|\geq n\varepsilon\right]\leq\frac{1}{n^2\varepsilon^2}Var\left[X_n\right]=\frac{\sum_{j=1}^{n}p_j(1-p_j)}{n^2\varepsilon^2}\leq\frac{\frac{1}{4}n}{n^2\varepsilon^2}=\frac{1}{4n\varepsilon^2}$

Así que:

$\text{lím}_{n\rightsquigarrow\infty}P\left[\left|\frac{X_n}{n}-\frac{1}{n}\sum_{j=1}^{n}p_j\right|>\varepsilon\right]\leq\text{lím}_{n\rightsquigarrow\infty}\frac{1}{4n\varepsilon^2}=0$

Las formas generales de estos resultados se expondrán en el segundo volumen de este libro.

9.9. Funciones generadoras

En esta sección vamos a introducir dos nuevos conceptos, el de función generadora de probabilidades y el de función generadora de momentos. El origen de ambos se remonta a Abraham de Moivre, quien, al resolver el problema de la determinación de la probabilidad de obtener una suma dada al lanzar n dados, introdujo un método que, al generalizarse, daría origen a una herramienta poderosa para el estudio de sumas de variables aleatorias independientes.

El problema que se planteó de Moivre es el siguiente: dados un número natural n y un número entero m, comprendido entre n y $6n$, ¿cuál es la probabilidad de obtener una suma igual a m al lanzar n dados?

Para resolver este problema, de Moivre consideró un número entero positivo t y n dados balanceados imaginarios, cada uno con $t+t^2+\cdots+t^6$ caras, de tal manera que t de ellas están marcadas con el número 1, t^2 con el número 2, etc. Consideró entonces el experimento aleatorio consistente en el lanzamiento de esos n dados imaginarios, para el cual hay $(t+t^2+\cdots+t^6)^n$ posibles resultados equiprobables. Entonces, el total de formas en que puede obtenerse una suma igual a m, al lanzar los dados imaginarios, está dado por el término que contiene a t^m en el desarrollo de $(t+t^2+\cdots+t^6)^n$. Además:

$(t+t^2+\cdots+t^6)^n=\left[t\frac{1-t^6}{1-t}\right]^n=t^n\left(1-t^6\right)^n\left(1-t\right)^{-n}$

$=t^n\left[\sum_{k=0}^{n}(-1)^k\binom{n}{k}t^{6k}\right]\left[\sum_{k=0}^{\infty}\binom{n+k-1}{n-1}t^k\right]$

Ahora bien, dado un número entero no negativo y, al desarrollar el producto

$\left[\sum_{k=0}^{n}(-1)^k\binom{n}{k}t^{6k}\right]\left[\sum_{k=0}^{\infty}\binom{n+k-1}{n-1}t^k\right]$

los términos que contienen t^y son aquellos que se obtienen de multiplicar un término del primer factor conteniendo t^{6k}, para alguna k tal que $y \geq 6k$, con un término del segundo factor conteniendo t^{y-6k}. De manera que:

$$\left[\sum_{k=0}^{n}(-1)^k\binom{n}{k}t^{6k}\right]\left[\sum_{k=0}^{\infty}\binom{n+k-1}{n-1}t^k\right] = \sum_{y=0}^{\infty}\sum_{\left\{k:0\leq k\leq \frac{y}{6}\right\}}(-1)^k\binom{n}{k}\binom{n+y-6k-1}{n-1}t^y$$

Así que:

$$(t+t^2+\cdots+t^6)^n = t^n\sum_{y=0}^{\infty}\sum_{\left\{k:0\leq k\leq \frac{y}{6}\right\}}(-1)^k\binom{n}{k}\binom{n+y-6k-1}{n-1}t^y$$

$$= \sum_{y=0}^{\infty}\sum_{\{k:0\leq 6k\leq y\}}(-1)^k\binom{n}{k}\binom{n+y-6k-1}{n-1}t^{n+y} = \sum_{x=n}^{\infty}\sum_{\{k:0\leq 6k\leq x-n\}}(-1)^k\binom{n}{k}\binom{x-6k-1}{n-1}t^x$$

Por lo tanto, si S es la suma de los números que se obtienen al lanzar los n dados imaginarios, se tiene:

$$P[S=m] = \frac{1}{(t+t^2+\cdots+t^6)^n}\sum_{\{k:0\leq 6k\leq m-n\}}(-1)^k\binom{n}{k}\binom{m-6k-1}{n-1}t^m$$

Finalmente, la solución al problema original se obtiene haciendo $t=1$, así que la probabilidad p_m de obtener una suma igual a m, al lanzar los n dados originales, está dada por:

$$p_m = \frac{1}{6^n}\sum_{\{k:0\leq 6k\leq m-n\}}(-1)^k\binom{n}{k}\binom{m-6k-1}{n-1}$$

Por ejemplo, la probabilidad de obtener una suma igual a 11, al lanzar 3 dados, está dada por:

$$p_3 = \frac{1}{6^3}\sum_{\{k:0\leq 6k\leq 8\}}(-1)^k\binom{3}{k}\binom{10-6k}{2} = \frac{1}{6^3}\sum_{k=0}^{1}(-1)^k\binom{3}{k}\binom{10-6k}{2} = \frac{1}{6^3}\left[\binom{10}{2}-3\binom{4}{2}\right] = \frac{1}{8}$$

9.9.1. Función generadora de probabilidades. El método utilizado por de Moivre en la solución del problema planteado en la introducción de esta sección motiva la siguiente definición:

DEFINICIÓN 9.99 (**Función generadora de probabilidades**). *Sea X una variable aleatoria discreta que admite como posibles valores únicamente enteros no negativos. A la función Φ_X definida por:*

$$\Phi_X(t) = \sum_{k=0}^{\infty}P[X=k]\,t^k = E\left[t^X\right]$$

para todos aquellos números reales t para los cuales la serie $\sum_{k=0}^{\infty}P[X=k]\,t^k$ converja absolutamente, se le llama la función generadora de probabilidades de X.

EJEMPLO 9.100. *Sea X una variable con distribución binomial negativa de parámetros r y p, entonces:*

$\Phi_X(t) = E\left[t^X\right] = p^r \sum_{x=0}^{\infty} t^x \binom{r+x-1}{x}(1-p)^x = p^r\left[1 - t(1-p)\right]^{-r} = \left[\frac{p}{1-t(1-p)}\right]^r$

para todos los números reales t tales que $|t(1-p)| < 1$, es decir, $|t| < \frac{1}{1-p}$.

▲

Obsérvese que, dada cualquier variable aleatoria discreta X que admita como posibles valores únicamente enteros no negativos, la serie:

$\sum_{k=0}^{\infty} P\left[X = k\right] t^k$

converge absolutamente para cualquier $t \in [-1, 1]$ pues en ese caso dicha serie está acotada por la serie convergente $\sum_{k=0}^{\infty} P\left[X = k\right] = 1$. Además, si la serie:

$\sum_{k=0}^{\infty} P\left[X = k\right] t_0^k$

converge para alguna $t_0 > 0$ y $t \in [-t_0, t_0]$, entonces:

$\sum_{k=0}^{\infty} P\left[X = k\right] |t|^k \leq \sum_{k=0}^{\infty} P\left[X = k\right] t_0^k$

así que la serie:

$\sum_{k=0}^{\infty} P\left[X = k\right] t^k$

converge absolutamente.

Por lo tanto, si:

$t_X = \sup\left\{t > 0 : \text{la serie } \sum_{k=0}^{\infty} P\left[X = k\right] t^k \text{ converge}\right\}$

entonces $t_X \geq 1$ y la serie $\sum_{k=0}^{\infty} P\left[X = k\right] t^k$ converge absolutamente para cualquier $t \in (-t_X, t_X)$ y diverge para cualquier $t \in (t_X, \infty) \cup (-\infty, -t_X)$.

De esta forma, la función generadora de probabilidades de X está definida ya sea en el intervalo $(-t_X, t_X)$ o bien en el intervalo $[-t_X, t_X]$.

El siguiente ejemplo muestra que la función generadora de probabilidades de una variable aleatoria X puede estar definida únicamente en el intervalo $[-1, 1]$.

Ejemplo 9.101. *Sea X una variable aleatoria con función de densidad dada por:*

$f_X(x) = \begin{cases} \frac{c}{x^2} & \text{si } x \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$

en donde c es una constante positiva tal que $\sum_{x=1}^{\infty} f_X(x) = 1$. Si $t > 1$, se tiene $\lim_{x\rightsquigarrow\infty} \frac{t^x}{x^2} = \infty$, así que la serie $\sum_{x=1}^{\infty} \frac{c}{x^2} t^x$ no converge. Por lo tanto, la función generadora de probabilidades de X está definida únicamente en el intervalo $[-1, 1]$.

▲

Las siguientes dos proposiciones establecen dos de las propiedades básicas de la función generadora de probabilidades.

Proposición 9.102. *Sean X y Y dos variables aleatorias discretas que admiten como posibles valores únicamente enteros no negativos y sean Φ_X y Φ_Y sus respectivas funciones generadoras de probabilidades. Supongamos además que $\Phi_X(t) = \Phi_Y(t)$ para cualquier $t \in [-1, 1]$, entonces X y Y tienen la misma función de densidad.*

Demostración

Por el teorema 25 del Apéndice, para cualquier $t \in (-1, 1)$, se tiene que tanto Φ_X como Φ_Y tienen derivadas de cualquier orden en t y, además:

$\Phi_X(t) = \sum_{k=0}^{\infty} P[X = k] t^k = P[X = 0] + \sum_{k=1}^{\infty} P[X = k] t^k$

$\Phi'_X(t) = \sum_{k=1}^{\infty} kP[X = k] t^{k-1} = P[X = 1] + \sum_{k=2}^{\infty} kP[X = k] t^{k-1}$

$\Phi''_X(t) = \sum_{k=2}^{\infty} k(k-1)P[X = k] t^{k-2} = 2!P[X = 2] + \sum_{k=3}^{\infty} k(k-1)P[X = k] t^{k-2}$

$\Phi'''_X(t) = \sum_{k=3}^{\infty} k(k-1)(k-2)P[X = k] t^{k-3}$

$= 3!P[X = 3] + \sum_{k=4}^{\infty} k(k-1)(k-2)P[X = k] t^{k-3}$

$\vdots$

$\Phi_X^{(n)}(t) = n!P[X = n] + \sum_{k=n+1}^{\infty} \frac{k!}{(k-n)!} P[X = k] t^{k-n}$

en donde $\Phi_X^{(n)}$ es la n-sima derivada de Φ_X.

De aquí se sigue que:

$\Phi_X^{(n)}(0) = n!P[X = n]$

De la misma manera, se tiene:

$\Phi_Y^{(n)}(0) = n!P[Y = n]$

Pero como $\Phi_X(t) = \Phi_Y(t)$ para cualquier $t \in [-1, 1]$, entonces $\Phi_X^{(n)}(0) = \Phi_Y^{(n)}(0)$ para cualquier $n \in \{0, 1, \ldots\}$, de lo cual se sigue el resultado.

■

PROPOSICIÓN 9.103. *Sean X y Y dos variables aleatorias discretas que admiten como posibles valores únicamente enteros no negativos y sean Φ_X y Φ_Y sus respectivas funciones generadoras de probabilidades. Supongamos además que X y Y son independientes, entonces $\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t)$ para cualquier t en la intersección de los dominios de definición de Φ_X y Φ_Y.*

Demostración

Como X y Y son independientes, se tiene $E\left[t^{X+Y}\right] = E[t^X]E[t^X]$, así que t pertenece al dominio de definición de Φ_{X+Y} si y sólo si t pertenece a la intersección de los dominios de definición de Φ_X y Φ_Y. Además, para una t con esa propiedad, se tiene:

$\Phi_{X+Y}(t) = E\left[t^{X+Y}\right] = E[t^X]E[t^X] = \Phi_X(t)\Phi_Y(t)$

∎

EJEMPLO 9.104. *Sean X y Y dos variable aleatorias independientes, ambas con distribución binomial negativa de parámetros r, p y s, p, respectivamente, entonces, para cualquier $t \in [0, 1]$, se tiene:*

$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t) = \left[\frac{p}{1-t(1-p)}\right]^r \left[\frac{p}{1-t(1-p)}\right]^s = \left[\frac{p}{1-t(1-p)}\right]^{r+s}$

Así que $X + Y$ tiene distribución binomial negativa con parámetros $r + s$ y p.

PROPOSICIÓN 9.105. *Sea X una variable aleatoria discreta que admite como posibles valores únicamente enteros no negativos y sea Φ_X su función generadora de probabilidades. Supongamos además que $t_X > 1$. Entonces todos los momentos de X son finitos y, para cualquier $n \in \mathbb{N}$, se tiene:*

$$E\left[X(X-1)\cdots(X-n+1)\right] = \Phi_X^{(n)}(1)$$

en donde $\Phi_X^{(n)}$ es la n-sima derivada de Φ_X.

Demostración

Por el teorema 25 del Apéndice, para cualquier $t \in (-t_X, t_X)$, Φ_X tiene derivadas de cualquier orden en t y, además:

$\Phi_X(t) = \sum_{k=0}^{\infty} P\left[X = k\right] t^k$

$\Phi_X'(t) = \sum_{k=1}^{\infty} kP\left[X = k\right] t^{k-1}$

$\Phi_X''(t) = \sum_{k=2}^{\infty} k(k-1)P\left[X = k\right] t^{k-2}$

$\Phi_X'''(t) = \sum_{k=3}^{\infty} k(k-1)(k-2)P\left[X = k\right] t^{k-3}$

$\vdots$

$\Phi_X^{(n)}(t) = \sum_{k=n}^{\infty} k(k-1)(k-2)(k-n+1)P[X=k]\, t^{k-n}$

De aquí se sigue que:

$\Phi_X^{(n)}(1) = \sum_{k=n}^{\infty} k(k-1)(k-2)(k-n+1)P[X=k] = E[X(X-1)\cdots(X-n+1)]$

Ejemplo 9.106. *Sea X una variable con distribución binomial negativa de parámetros r y p, entonces:*

$\Phi_X(t) = p^r \left[\frac{1}{1-t(1-p)}\right]^r$

$\Phi_X'(t) = rp^r \left[\frac{1}{1-t(1-p)}\right]^{r-1} \frac{1-p}{[1-t(1-p)]^2} = r(1-p)p^r \left[\frac{1}{1-t(1-p)}\right]^{r+1}$

$\Phi_X''(t) = r(1-p)p^r(r+1) \left[\frac{1}{1-t(1-p)}\right]^{r} \frac{1-p}{[1-t(1-p)]^2} = r(r+1)(1-p)^2 p^r \left[\frac{1}{1-t(1-p)}\right]^{r+2}$

Así que:

$E[X] = \Phi_X'(1) = r\frac{1-p}{p}$

$E[X(X-1)] = \Phi_X''(1) = r(r+1)\left(\frac{1-p}{p}\right)^2$

$E[X^2] = E[X(X-1)] + E[X] = r(r+1)\left(\frac{1-p}{p}\right)^2 + r\frac{1-p}{p}$

$Var(X) = E[X^2] - (E[X])^2 = r(r+1)\left(\frac{1-p}{p}\right)^2 + r\frac{1-p}{p} - r^2\left(\frac{1-p}{p}\right)^2$

$= r\left(\frac{1-p}{p}\right)^2 + r\frac{1-p}{p} = r\frac{1-p}{p}\left(1 + \frac{1-p}{p}\right) = r\frac{1-p}{p^2}$

▲

La proposición 9.105 permite calcular la esperanza de una variable aleatoria, con valores enteros no negativos, a partir de su función generadora de probabilidades para el caso en que esta última esté definida en una vecindad de $t = 1$. Sin embargo, como muestra el ejemplo 9.101, la función generadora pudiera estar definida únicamente en el intervalo $[-1, 1]$. La siguiente proposición permite el cálculo de la esperanza, cuando es finita, en esta situación.

Proposición 9.107. *Sea X una variable aleatoria discreta de esperanza finita que admite como posibles valores únicamente enteros no negativos y sea Φ_X su función generadora de probabilidades. Entonces:*

$$E[X] = \text{lím}_{t \rightsquigarrow 1-} \Phi'_X(t)$$

Demostración

Por el teorema 25 del Apéndice, la serie $\sum_{k=1}^{\infty} kP[X = k] t^{k-1}$ converge absolutamente para $t \in [0, 1)$ y, como X tiene esperanza finita, también se tiene la convergencia absoluta para $x = 1$. Así que, Por el teorema 26 del Apéndice, si definimos $f : [0, 1] \mapsto \mathbb{R}$ mediante la relación $f(t) = \sum_{k=1}^{\infty} kP[X = k] t^{k-1}$, entonces $\sum_{k=1}^{\infty} kP[X = k] = \text{lím}_{x \rightsquigarrow 1-} f(x)$.

Además, también por el teorema 25 del Apéndice, para cualquier $t \in [0, 1)$, Φ_X es derivable en t y $\Phi'_X(t) = \sum_{k=1}^{\infty} kP[X = k] t^{k-1} = f(t)$.

Por lo tanto:

$E[X] = \sum_{k=1}^{\infty} kP[X = k] = \text{lím}_{t \rightsquigarrow 1-} f(t) = \text{lím}_{t \rightsquigarrow 1-} \Phi'_X(t)$

■

Ejemplo 9.108 (**Duración esperada del juego en el problema de la ruina del jugador**). *Dos jugadores, A y B, los cuales poseen, respectivamente, a y b fichas, juegan partidas consecutivas en cada una de las cuales las probabilidades de que A y B ganen son, respectivamente, p y $q = 1 - p$ y de tal manera que en cada una de ellas el perdedor dará una de sus fichas al vencedor. Si el juego se termina en el momento en que alguno de los dos jugadores llegue a poseer la totalidad de fichas, ¿cuál es la duración esperada del juego?*

Solución

El experimento aleatorio $\mathcal{E}$ asociado con este problema puede verse como la realización de una sucesión de ensayos de Bernoulli $B_1, B_2, \ldots$, en donde B_i es éxito si el jugador A gana la i-ésima partida.

Sea $t = a + b - 1$ y, para cada $n \in \mathbb{N}$, definamos el evento:

A_n: Alguno de los dos jugadores se arruina en alguno de los primeros n ensayos.

Demostremos primero que $P(A_{kt}^c) \leq (1 - p^t)^k$ para cualquier $k \in \mathbb{N}$. La demostración será por inducción:

Si A gana las primeras b partidas, entonces ocurre A_t, así que $P(A_t) \geq p^b \geq p^t$. Por lo tanto $P(A_t^c) \leq 1 - p^t$.

Supongamos ahora que la desigualdad $P(A_{kt}^c) \leq (1 - p^t)^k$ se cumple para $k = m \in \mathbb{N}$, entonces:

$P\left(A^c_{(m+1)t}\right) = P\left(A^c_{(m+1)t} \cap A^c_{mt}\right) = P\left(A^c_{mt}\right) - P\left(A_{(m+1)t} \cap A^c_{mt}\right)$

$= P\left(A^c_{mt}\right) - P\left(A_{(m+1)t} \mid A^c_{mt}\right) P\left(A^c_{mt}\right)$

Pero, dado que A^c_{mt} ocurre, Si A gana consecutivamente las siguientes partidas hasta quedarse con todas las fichas, entonces ocurre $A_{(m+1)t}$, así que $P\left(A_{(m+1)t} \mid A^c_{mt}\right) \geq p^t$. Por lo tanto:

$P\left(A^c_{(m+1)t}\right) = P\left(A^c_{mt}\right) - P\left(A_{(m+1)t} \mid A^c_{mt}\right) P\left(A^c_{mt}\right)$

$\leq P\left(A^c_{mt}\right) - p^t P\left(A^c_{mt}\right) = \left(1 - p^t\right) P\left(A^c_{mt}\right) \leq \left(1 - p^t\right)\left(1 - p^t\right)^m = \left(1 - p^t\right)^{m+1}$

Sea ahora T el número de ensayo en el cual alguno de los dos jugadores queda arruinado. Entonces, para cualquier $k \in \mathbb{N}$, se tiene:

$P\left[T > tk\right] = P\left(A^c_{tk}\right) \leq \left(1 - p^t\right)^k$

Así que:

$E\left[T\right] = \sum_{n=1}^{\infty} P\left[T \geq n\right] = 1 + \sum_{n=1}^{\infty} P\left[T > n\right]$

$= 1 + P\left[T > 1\right] + \cdots + P\left[T > t\right] + P\left[T > t+1\right] + \cdots + P\left[T > 2t\right] + \cdots =$

$\leq 1 + tP\left[T > 1\right] + tP\left[T > t\right] + tP\left[T > 2t\right] + \cdots$

$\leq 1 + t + t\sum_{k=1}^{\infty} P\left[T > tk\right] \leq 1 + t\sum_{k=0}^{\infty}\left(1 - p^t\right)^k = 1 + \frac{t}{p^t} < \infty$

Sea $T_0 = T_{a+b} = 0$ y, para $z \in \{1, \ldots, a+b-1\}$, sea T_z el número de ensayo en el cual alguno de los dos jugadores queda arruinado a partir del momento en que el jugador A tiene z fichas. Sea además, para $z \in \{0, \ldots, a+b\}$, Φ_z la función generadora de probabilidades de T_z.

Se tiene $P\left[T_0 = 0\right] = P\left[T_{a+b} = 0\right] = 1$, así que $\Phi_0(t) = \Phi_{a+b}(t) \equiv 1$.

Para $z \in \{1, \ldots, a+b-1\}$, se tiene $P\left[T_z = 0\right] = 0$ y, si $n \in \mathbb{N}$:

$P\left[T_z = n\right] = pP\left[T_{z+1} = n-1\right] + qP\left[T_{z-1} = n-1\right]$

Así que:

$\Phi_z(t) = E\left[t^{T_z}\right] = \sum_{k=1}^{\infty} t^k P\left[T_z = k\right]$

$= p\sum_{k=1}^{\infty} t^k P\left[T_{z+1} = k-1\right] + q\sum_{k=1}^{\infty} t^k P\left[T_{z-1} = k-1\right]$

$= pt\sum_{k=1}^{\infty} t^{k-1} P\left[T_{z+1} = k-1\right] + qt\sum_{k=1}^{\infty} t^{k-1} P\left[T_{z-1} = k-1\right]$

$= pt \sum_{k=0}^{\infty} t^k P[T_{z+1} = k] + qt \sum_{k=0}^{\infty} t^k P[T_{z-1} = k]$

$= pt\Phi_{z+1}(t) + qt\Phi_{z-1}(t)$

Por lo tanto, para cualquier $t \in (0,1)$*, se tiene:*

$\Phi_z'(t) = pt\Phi_{z+1}'(t) + qt\Phi_{z-1}'(t) + p\Phi_{z+1}(t) + q\Phi_{z-1}(t)$

De manera que, tomando límites cuando t *tiende a* 1 *por la izquierda y utilizando la proposición 9.107, se tiene:*

$E[T_z] = pE[T_{z+1}] + qE[T_{z-1}] + 1$

Para encontrar el valor de $E[T_z]$ *tenemos que resolver entonces la ecuación en diferencias:*

$m_z = pm_{z+1} + qm_{z-1} + 1$ *para* $z \in \{1, \ldots, a+b-1\}$

Una solución particular de esta ecuación está dada por:

$\alpha_z = \begin{cases} -z^2 & si\ p = q \\ \frac{z}{q-p} & si\ p \neq q \end{cases}$

Si $\{\beta_z\}_{z\in\{0,1,\ldots,a+b\}}$ *es otra solución, entonces* $\gamma_z = \beta_z - \alpha_z$ *es solución de:*

$\gamma_z = p\gamma_{z+1} + q\gamma_{z-1}$ *para* $z \in \{1, \ldots, a+b-1\}$

Para encontrar la solución general de esta ecuación, la escribiremos en la siguiente forma equivalente, la cual se obtiene escribiendo $\gamma_z = \gamma_z(p+q)$.

$q(\gamma_z - \gamma_{z-1}) = p(\gamma_{z+1} - \gamma_z)$

Supongamos primero que $p = q$*, entonces se tiene:*

$\gamma_1 - \gamma_0 = \gamma_2 - \gamma_1 = \cdots = \gamma_{a+b} - \gamma_{a+b-1} = c$

en donde c *es una constante. Así que, para* $z \in \{1, \ldots, a+b\}$*:*

$\gamma_z - \gamma_0 = \sum_{k=1}^{z} (\gamma_k - \gamma_{k-1}) = cz$

Por lo tanto, $\gamma_z = \gamma_0 + cz$

Supongamos ahora $p \neq q$*. En este caso tenemos, para* $z \in \{1, \ldots, a+b-1\}$*:*

$q^z \prod_{k=1}^{z} (\gamma_k - \gamma_{k-1}) = p^z \prod_{k=1}^{z} (\gamma_{k+1} - \gamma_k)$

Así que:

$\gamma_{z+1} - \gamma_z = \left(\frac{q}{p}\right)^z (\gamma_1 - \gamma_0)$

y entonces, para $z \in \{1, \ldots, a+b\}$*:*

$\gamma_z - \gamma_0 = \sum_{k=0}^{z-1} \left(\gamma_{k+1} - \gamma_k\right) = (\gamma_1 - \gamma_0) \sum_{k=0}^{z-1} \left(\frac{q}{p}\right)^k = (\gamma_1 - \gamma_0) \frac{1-\left(\frac{q}{p}\right)^z}{1-\frac{q}{p}}$

Por lo tanto:

$\gamma_z = 1 + (\gamma_1 - 1) \frac{1-\left(\frac{q}{p}\right)^z}{1-\frac{q}{p}} = A + B\left(\frac{q}{p}\right)^z$

en donde A *y* B *son constantes.*

Así que:

$$\gamma_z = \begin{cases} \gamma_0 + cz & si\ p = q \\ A + B\left(\frac{q}{p}\right)^z & si\ p \neq q \end{cases}$$

Por lo tanto:

$$\beta_z = \alpha_z + \gamma_z = \begin{cases} -z^2 + \gamma_0 + cz & si\ p = q \\ \frac{z}{q-p} + A + B\left(\frac{q}{p}\right)^z & si\ p \neq q \end{cases}$$

es la solución general de la ecuación original.

Así que:

$$m_z = \begin{cases} -z^2 + \gamma_0 + cz & si\ p = q \\ \frac{z}{q-p} + A + B\left(\frac{q}{p}\right)^z & si\ p \neq q \end{cases}$$

Ahora bien, como $m_0 = m_{a+b} = 0$*, se tiene:*

$$0 = \begin{cases} \gamma_0 & si\ p = q \\ A + B & si\ p \neq q \end{cases}$$

$$0 = \begin{cases} -(a+b)^2 + \gamma_0 + c(a+b) & si\ p = q \\ \frac{a+b}{q-p} + A + B\left(\frac{q}{p}\right)^{a+b} & si\ p \neq q \end{cases}$$

Así que:

$\gamma_0 = 0$*;* $c = a + b$*;* $A = -\frac{a+b}{q-p}\frac{1}{1-\left(\frac{q}{p}\right)^{a+b}}$*;* $B = -A$

Por lo tanto:

$$m_z = \begin{cases} -z^2+(a+b)z & si\ p=q \\ \frac{z}{q-p} + A\left[1-\left(\frac{q}{p}\right)^z\right] & si\ p\neq q \end{cases} = \begin{cases} (a+b-z)z & si\ p=q \\ \frac{z}{q-p} - \frac{a+b}{q-p}\frac{1-\left(\frac{q}{p}\right)^z}{1-\left(\frac{q}{p}\right)^{a+b}} & si\ p\neq q \end{cases}$$

Para el problema planteado inicialmente, en el cual el jugador A comienza con a fichas y el jugador B con b fichas, se tiene:

$$E\left[T\right] = \begin{cases} ab & si\ p=q \\ \frac{a}{q-p} - \frac{a+b}{q-p}\frac{1-\left(\frac{q}{p}\right)^a}{1-\left(\frac{q}{p}\right)^{a+b}} & si\ p\neq q \end{cases}$$

9.9.2. Función generadora de momentos.

DEFINICIÓN 9.109 (**Función generadora de momentos**). *Sea X una variable aleatoria. A la función M_X definida por:*

$$M_X(t) = E[e^{tX}]$$

para todos aquellos números reales t para los cuales e^{tX} tenga esperanza finita, se le llama la función generadora de momentos de X

EJEMPLO 9.110. *Sea X una variable con distribución gama de parámetros α y λ, entonces:*

$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx}\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}dx = \int_0^\infty \frac{\lambda^\alpha x^{\alpha-1}e^{-(\lambda-t)x}}{\Gamma(\alpha)}dx$

$= \frac{\lambda^\alpha}{(\lambda-t)^\alpha}\int_0^\infty \frac{(\lambda-t)^\alpha x^{\alpha-1}e^{-(\lambda-t)x}}{\Gamma(\alpha)}dx = \left(\frac{\lambda}{\lambda-t}\right)^\alpha$

para todos los números reales t tales que $\lambda - t > 0$, es decir, $t < \lambda$.

EJEMPLO 9.111. *Sea X una variable con distribución normal estándar, entonces:*

$M_X(t) = E[e^{tX}] = \int_{-\infty}^\infty e^{tx}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{tx}e^{-\frac{1}{2}(x^2-2tx)}dx$

$= \frac{1}{\sqrt{2\pi}}e^{\frac{1}{2}t^2}\int_{-\infty}^\infty e^{-\frac{1}{2}(x-t)^2}dx = e^{\frac{1}{2}t^2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2}y^2}dy = e^{\frac{1}{2}t^2}$

para cualquier número real t.

Ahora, si Y es una variable con distribución normal de parámetros μ y σ^2, entonces $X = \frac{Y-\mu}{\sigma^2}$ tiene distribución normal estándar, así que:

$M_Y(t) = E[e^{tY}] = E[e^{t(\mu+\sigma^2 X)}] = e^{\mu t}E[e^{\sigma^2 tX}]$

$= e^{\mu t}M_X(\sigma^2 t) = e^{\mu t}e^{\frac{1}{2}\sigma^2 t^2} = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}$

para cualquier número real t.

▲

Obsérvese que e^{tX} tiene esperanza finita por lo menos para $t = 0$. El siguiente ejemplo muestra que la función generadora de momentos de una variable aleatoria X puede estar definida únicamente en $t = 0$.

EJEMPLO 9.112. *Sea* X *una variable aleatoria con distribución Cauchy, es decir, con función de densidad dada por* $f_X(x) = \frac{1}{\pi(1+x^2)}$ *para cualquier* $x \in \mathbb{R}$.

Si $t > 0$, *se tiene:*

$\lim_{x \rightsquigarrow \infty} \frac{e^{tx}}{1+x^2} = \infty$

Así que:

$\int_0^\infty \frac{e^{tx}}{1+x^2} dx = \infty$

Si $t < 0$, *se tiene:*

$\lim_{x \rightsquigarrow -\infty} \frac{e^{tx}}{1+x^2} = \infty$

Así que:

$\int_{-\infty}^0 \frac{e^{tx}}{1+x^2} dx = \infty$

Por lo tanto, para cualquier $t \neq 0$, *se tiene:*

$M_X(t) = E[e^{tX}] = \frac{1}{\pi} \int_0^\infty \frac{e^{tx}}{1+x^2} dx + \frac{1}{\pi} \int_{-\infty}^0 \frac{e^{tx}}{1+x^2} dx = \infty$

EJEMPLO 9.113. *Sea* X *una variable aleatoria con función de densidad dada por* $f_X(x) = \frac{1}{4} e^{-\sqrt{|x|}}$ *para cualquier* $x \in \mathbb{R}$.

Si $t > 0$, *se tiene* $\lim_{x \rightsquigarrow \infty}(tx - \sqrt{x}) = \lim_{x \rightsquigarrow \infty}(t\sqrt{x} - 1)\sqrt{x} = \infty$, *así que* $\int_0^\infty e^{tx - \sqrt{x}} dx = \infty$.

Si $t < 0$, *se tiene* $\lim_{x \rightsquigarrow \infty}(-tx - \sqrt{x}) = \lim_{x \rightsquigarrow \infty}(-t\sqrt{x} - 1)\sqrt{x} = \infty$,, *así que* $\int_0^\infty e^{-tx - \sqrt{x}} dx = \infty$.

Por lo tanto, para cualquier $t \neq 0$, *se tiene:*

$M_X(t) = E[e^{tX}] = \int_{-\infty}^\infty e^{tx} e^{-\sqrt{|x|}} dx = \int_0^\infty e^{tx - \sqrt{x}} dx + \int_{-\infty}^0 e^{tx - \sqrt{-x}} dx$

$= \int_0^\infty e^{tx - \sqrt{x}} dx + \int_0^\infty e^{-tx - \sqrt{x}} dx = \infty$

▲

Por otra parte, definiendo $Z = e^{tX}$, se tiene, para $z > 0$:

$F_Z(z) = P\left[e^{tX} \leq z\right] = P\left[tX \leq \ln z\right]$

$= \begin{cases} P\left[X \leq \frac{1}{t}\ln z\right] & \text{si } t > 0 \\ P\left[X \geq \frac{1}{t}\ln z\right] & \text{si } t < 0 \end{cases} = \begin{cases} F_X(\frac{1}{t}\ln z) & \text{si } t > 0 \\ P\left[X \geq \frac{1}{t}\ln z\right] & \text{si } t < 0 \end{cases}$

Así que:

$M_X(t) = E[e^{tX}] = E[Z] = \int_0^\infty \left[1 - F_Z(z)\right] dz$

$= \begin{cases} \int_0^\infty \left[1 - F_X(\frac{1}{t}\ln z)\right] dz & \text{si } t > 0 \\ \int_0^\infty F_X(\frac{1}{t}\ln z) dz & \text{si } t < 0 \end{cases}$

$= \begin{cases} t\int_{-\infty}^\infty \left[1 - F_X(x)\right] e^{tx} dx & \text{si } t > 0 \\ -t\int_{-\infty}^\infty F_X(x) e^{tx} dx & \text{si } t < 0 \end{cases}$

$= \begin{cases} t\int_0^\infty \left[1 - F_X(x)\right] e^{tx} dx + t\int_{-\infty}^0 \left[1 - F_X(x)\right] e^{tx} dx & \text{si } t > 0 \\ -t\int_{-\infty}^0 F_X(x) e^{tx} dx - t\int_0^\infty F_X(x) e^{tx} dx & \text{si } t < 0 \end{cases}$

$= 1 + t\int_0^\infty \left[1 - F_X(x)\right] e^{tx} dx - t\int_{-\infty}^0 F_X(x) e^{tx} dx$

Para $t > 0$, la integral $\int_{-\infty}^0 F_X(x) e^{tx} dx$ siempre es finita, por lo tanto, $M_X(t) < \infty$ si y sólo si $\int_0^\infty \left[1 - F_X(x)\right] e^{tx} dx < \infty$; mientras que si $t < 0$, la integral $\int_0^\infty \left[1 - F_X(x)\right] e^{tx} dx$ siempre es finita, por lo tanto $M_X(t) < \infty$ si y sólo si $\int_{-\infty}^0 F_X(x) e^{tx} dx < \infty$.

Si $\int_0^\infty \left[1 - F_X(x)\right] e^{t_0 x} dx < \infty$ para alguna $t_0 > 0$ y $0 < t < t_0$, entonces $e^{tx} < e^{t_0 x}$ para $x > 0$, así que $\int_0^\infty \left[1 - F_X(x)\right] e^{tx} dx < \infty$. Por lo tanto $M_X(t) < \infty$ para cualquier $t \in [0, t_0]$.

Si $\int_{-\infty}^0 F_X(x) e^{s_0 x} dx < \infty$ para alguna $s_0 < 0$ y $s_0 < t < 0$, entonces $e^{tx} < e^{s_0 x}$ para $x < 0$, así que $\int_{-\infty}^0 F_X(x) e^{t_0 x} dx < \infty$. Por lo tanto $M_X(t) < \infty$ para cualquier $t \in [s_0, 0]$.

De esta forma, la función generadora de momentos de X está definida en un intervalo de extremos s_0 y t_0.

PROPOSICIÓN 9.114. *Sea X una variable aleatoria y M_X su función generadora de momentos. Supongamos además que M_X está definida en un intervalo $(-t_0, t_0)$, en donde $t_0 > 0$. Entonces todos los momentos de X son finitos y:*

$$M_X(t) = \sum_{k=0}^\infty \frac{t^k}{k!} E\left[X^k\right]$$

para cualquier $t \in (-t_0, t_0)$.

Demostración

Definiendo $Z = |X|^k$, se tiene, para $z > 0$:

$$F_Z(z) = P\left[|X|^k \leq z\right] = P\left[|X| \leq z^{\frac{1}{k}}\right] = P\left[-z^{\frac{1}{k}} \leq X \leq z^{\frac{1}{k}}\right] = F_X(z^{\frac{1}{k}}) - P\left[X < z^{\frac{1}{k}}\right]$$

Así que:

$$E[|X|^k] = \int_0^\infty [1 - F_Z(z)]\, dz = \int_0^\infty \left[1 - F_X(z^{\frac{1}{k}}) + F_X(-z^{\frac{1}{k}})\right] dz$$

$$= k \int_0^\infty [1 - F_X(x) + F_X(-x)]\, x^{k-1} dx$$

$$= k \int_0^\infty [1 - F_X(x)]\, x^{k-1} dx + k \int_0^\infty F_X(-x) x^{k-1} dx$$

$$= k \int_0^\infty [1 - F_X(x)]\, x^{k-1} dx + k \int_0^\infty F_X(-x) x^{k-1} dx$$

$$= k \int_0^\infty [1 - F_X(x)]\, x^{k-1} dx + k \int_{-\infty}^0 F_X(x)\, |x|^{k-1}\, dx$$

Ahora bien, recordemos que se tiene:

$$M_X(t) = 1 + t \int_0^\infty [1 - F_X(x)]\, e^{tx} dx - t \int_{-\infty}^0 F_X(x) e^{tx} dx$$

y que, por hipótesis, M_X es finita en en un intervalo $(-t_0, t_0)$, en donde $t_0 > 0$, así que $\int_0^\infty [1 - F_X(x)]\, e^{t_0 x} dx < \infty$ y $\int_{-\infty}^0 F_X(x) e^{-t_0 x} dx < \infty$. Además, $e^{tx} = \sum_{k=0}^\infty \frac{t^k}{k!} x^k$, así que $x^k \leq \frac{k!}{t^k} e^{t_0 x}$ para cualquier $x > 0$ y cualquier k y $|x|^k \leq \frac{k!}{|t|^k} e^{s_0 x}$ para cualquier $x < 0$ y cualquier k.

Por lo tanto:

$$\int_0^\infty [1 - F_X(x)]\, x^{k-1} dx < \infty$$

$$\int_{-\infty}^0 F_X(x)\, |x|^{k-1}\, dx < \infty$$

Así que $E[|X|^k] < \infty$.

Para k impar, definiendo $Y = X^k$, se tiene:

$$F_Y(y) = P\left[X^k \leq y\right] = P\left[X \leq y^{\frac{1}{k}}\right] = F_X(y^{\frac{1}{k}})$$

Así que:

$$E[X^k] = \int_0^\infty [1 - F_Y(y)]\, dy - \int_{-\infty}^0 F_Y(y) dy$$

$= \int_0^\infty \left[1 - F_X(y^{\frac{1}{k}})\right] dy - \int_{-\infty}^0 F_X(y^{\frac{1}{k}}) dy$

$= k \int_0^\infty [1 - F_X(x)] \, x^{k-1} dx - k \int_{-\infty}^0 F_X(x) x^{k-1} dx$

$= k \int_0^\infty [1 - F_X(x)] \, x^{k-1} dx - k \int_{-\infty}^0 F_X(x) x^{k-1} dx$

Por otra parte, si k es par:

$E[X^k] = E[|X|^k] = k \int_0^\infty [1 - F_X(x)] \, x^{k-1} dx + k \int_{-\infty}^0 F_X(x) \, |x|^{k-1} \, dx$

$= k \int_0^\infty [1 - F_X(x)] \, x^{k-1} dx - k \int_{-\infty}^0 F_X(x) x^{k-1} dx$

Así que, en cualquier caso, para $k > 0$:

$E[X^k] = k \int_0^\infty [1 - F_X(x)] \, x^{k-1} dx - k \int_{-\infty}^0 F_X(x) x^{k-1} dx$

Por lo tanto:

$\sum_{k=0}^n \frac{t^k}{k!} E\left[X^k\right] = 1 + \sum_{k=1}^n \frac{t^k}{k!} E\left[X^k\right]$

$= 1 + \sum_{k=1}^n \frac{t^k}{k!} k \int_0^\infty [1 - F_X(x)] \, x^{k-1} dx - \sum_{k=1}^n \frac{t^k}{k!} k \int_{-\infty}^0 F_X(x) x^{k-1} dx$

$= 1 + t \int_0^\infty [1 - F_X(x)] \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k dx - t \int_{-\infty}^0 F_X(x) \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k dx$

Pero como $e^{tx} = \lim_{n \rightsquigarrow \infty} \sum_{k=0}^n \frac{t^k}{k!} x^k$, se tiene:

Si $t \in [0, t_0)$, la sucesión de funciones $x \mapsto [1 - F_X(x)] \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k$, para $x > 0$, es monótona no decreciente, así que, por el teorema de la convergencia monótona, se tiene:

$\lim_{n \rightsquigarrow \infty} \int_0^\infty [1 - F_X(x)] \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k dx$

$= \int_0^\infty [1 - F_X(x)] \lim_{n \rightsquigarrow \infty} \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k dx = \int_0^\infty [1 - F_X(x)] \, e^{tx} dx$

Para $t \in [0, t_0)$ y $x < 0$, se tiene:

$\left| \sum_{k=0}^n \frac{t^k}{k!} x^k \right| \leq \sum_{k=0}^n \frac{t^k}{k!} |x|^k \leq \sum_{k=0}^\infty \frac{t^k}{k!} |x|^k = e^{t|x|} = e^{-tx}$

y, además:

$\int_{-\infty}^0 F_X(x) e^{-tx} dx < \infty \leq \int_{-\infty}^0 F_X(x) e^{-t_0 x} dx < \infty$

Así que, por el teorema de la convergencia dominada, se tiene:

$\lim_{n \rightsquigarrow \infty} \int_{-\infty}^{0} F_X(x) \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k dx$

$= \int_{-\infty}^{0} F_X(x) \lim_{n \rightsquigarrow \infty} \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k dx = \int_{-\infty}^{0} F_X(x) e^{tx} dx$

Si $t \in (-t_0, 0)$, la sucesión de funciones $x \mapsto F_X(x) \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k$, para $x < 0$, es monótona no decreciente, así que, por el teorema de la convergencia monótona, se tiene:

$\lim_{n \rightsquigarrow \infty} \int_{-\infty}^{0} F_X(x) \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k dx$

$= \int_{-\infty}^{0} F_X(x) \lim_{n \rightsquigarrow \infty} \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k dx = \int_{-\infty}^{0} F_X(x) e^{tx} dx$

Si $t \in (-t_0, 0)$ y $x > 0$, se tiene:

$\left| \sum_{k=0}^{n} \frac{t^k}{k!} x^k \right| \leq \sum_{k=0}^{n} \frac{|t|^k}{k!} x^k \leq \sum_{k=0}^{\infty} \frac{|t|^k}{k!} x^k = e^{|t|x} = e^{-tx}$

y, además:

$\int_0^\infty [1 - F_X(x)]\, e^{-tx} dx < \infty \leq \int_0^\infty [1 - F_X(x)]\, e^{t_0 x} dx < \infty$

Así que, por el teorema de la convergencia dominada, se tiene:

$\lim_{n \rightsquigarrow \infty} \int_0^\infty [1 - F_X(x)] \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k dx$

$= \int_0^\infty [1 - F_X(x)] \lim_{n \rightsquigarrow \infty} \sum_{k=0}^{n-1} \frac{t^k}{k!} x^k dx = \int_0^\infty [1 - F_X(x)]\, e^{tx} dx$

Por lo tanto:

$\sum_{k=0}^{\infty} \frac{t^k}{k!} E\left[X^k\right] = \lim_{n \rightsquigarrow \infty} \sum_{k=0}^{n} \frac{t^k}{k!} E\left[X^k\right]$

$= 1 + t \int_0^\infty [1 - F_X(x)]\, e^{tx} dx - t \int_{-\infty}^{0} F_X(x) e^{tx} dx = M_X(t)$

■

COROLARIO 9.115. *Sea X una variable aleatoria y sea M_X su función generadora de momentos. Supongamos además que M_X está definida en una vecindad de 0. Entonces, para cualquier $n \in \mathbb{N}$, se tiene:*

$$E[X^n] = M_X^{(n)}(0)$$

en donde $M_X^{(n)}$ denota la enésima derivada de M_X.

EJEMPLO 9.116. *Sea X una variable aleatoria con distribución uniforme en el intervalo $[a, b]$, entonces:*

$M_X(t) = E[e^{tX}] = \frac{1}{b-a}\int_a^b e^{tx}dx = \begin{cases} \frac{1}{t(b-a)}\left(e^{tb} - e^{ta}\right) & si\ t \neq 0 \\ 1 & si\ t = 0 \end{cases}$

$M'_X(t) = \frac{1}{t(b-a)}\left(be^{tb} - ae^{ta}\right) - \frac{1}{t^2(b-a)}\left(e^{tb} - e^{ta}\right)$

$M''_X(t) = \frac{1}{t(b-a)}\left(b^2e^{tb} - a^2e^{ta}\right) - \frac{2}{t^2(b-a)}\left(be^{tb} - ae^{ta}\right) + \frac{2}{t^3(b-a)}\left(e^{tb} - e^{ta}\right)$

Así que:

$E[X] = \text{lím}_{t\rightsquigarrow 0}\left[\frac{1}{t(b-a)}\left(be^{tb} - ae^{ta}\right) - \frac{1}{t^2(b-a)}\left(e^{tb} - e^{ta}\right)\right] = \frac{1}{2}(a+b)$

$E[X^2] = \text{lím}_{t\rightsquigarrow 0}\left[\frac{1}{t(b-a)}\left(b^2e^{tb} - a^2e^{ta}\right) - \frac{2}{t^2(b-a)}\left(be^{tb} - ae^{ta}\right) + \frac{2}{t^3(b-a)}\left(e^{tb} - e^{ta}\right)\right]$

$= \frac{1}{3}(a^2 + ab + b^2)$

$Var(X) = E[X^2] - (E[X])^2 = \frac{1}{3}(a^2 + ab + b^2) - \frac{1}{4}(a+b)^2 = \frac{1}{12}(b-a)^2$

▲

Las siguientes dos proposiciones establecen dos de las propiedades básicas de la función generadora de momentos. La demostración de la primera de ellas requiere de herramientas no desarrolladas en este libro. Puede consultarse una demostración en Billingsley, P., Probability and Measure, John Wiley, 1979.

PROPOSICIÓN 9.117. *Sean X y Y dos variables aleatorias y sean M_X y M_Y sus respectivas funciones generadoras de momentos. Supongamos además que M_X y M_Y están definidas y son iguales en una vecindad de* 0, *entonces X y Y tienen la misma distribución.*

PROPOSICIÓN 9.118. *Sean X y Y dos variables aleatorias independientes y sean M_X y M_Y sus respectivas funciones generadoras de momentos. Entonces $M_{X+Y}(t) = M_X(t)M_Y(t)$ para cualquier t en la intersección de los dominios de definición de M_X y M_Y.*

Demostración

Como X y Y son independientes, se tiene $E[e^{t(X+Y)}] = E[e^{tX)}]E[e^{tY}]$, así que t pertenece al dominio de definición de M_{X+Y} si y sólo si t pertenece a la intersección de los dominios de definición de M_X y M_Y. Además, para una t con esa propiedad, se tiene:

$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX)}]E[e^{tY}] = M_X(t)M_Y(t)$

■

EJEMPLO 9.119. *Sean X y Y dos variables aleatorias independientes, ambas con distribución gama de parámetros α_1, λ y α_2, λ, respectivamente, entonces, para cualquier $t < \lambda$, se tiene:*

$$M_{X+Y}(t) = M_X(t) M_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1} \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_2} = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1 + \alpha_2}$$

Así que $X + Y$ tiene distribución gama con parámetros $\alpha_1 + \alpha_2$ y λ.

EJERCICIOS

EJERCICIO 9.1. *Utilice el método del ejemplo 9.5 para calcular la esperanza de una variable aleatoria X con distribución Poisson de parámetro λ.*

EJERCICIO 9.2. *Utilice el método del ejemplo 9.6 para calcular la esperanza de una variable aleatoria X con distribución a) binomial de parámetros n y p y b) binomial negativa de parámetros r y p.*

EJERCICIO 9.3. *Sea X una variable aleatoria con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{2x}{N(N+1)} & \text{si } x \in \{1, \ldots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde N es un entero positivo. Encuentre $E[X]$.

EJERCICIO 9.4. *Sea X una variable aleatoria con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{1}{x!} & \text{si } x \in \{2, 3, \ldots\} \\ 3 - e & \text{si } x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre $E[X]$.

EJERCICIO 9.5. *Un juego popular en Inglaterra consiste en el lanzamiento de 3 dados. Un jugador puede apostar a cualquiera de los números enteros entre 1 y 6. Si se obtiene ese número en exactamente uno de los 3 lanzamientos, se gana lo mismo que se apostó; si se obtiene en exactamente dos de los 3 lanzamientos, se gana el doble de la apuesta y si se obtiene en los 3 lanzamientos se gana el triple de la apuesta. Por otro lado, si no se obtiene el número en ninguno de los 3 lanzamientos, el jugador pierde la apuesta. Encuentre la esperanza de la ganancia (o pérdida) de un jugador que apuesta un peso en dicho juego.*

EJERCICIO 9.6. *Un juego consiste en el lanzamiento de n dados. Un jugador elige un número entero entre 1 y 6. Por cada dado con el cual se obtenga ese número el* *jugador gana* 1 *peso. Por otro lado, si no se obtiene el número en ninguno de los* n *lanzamientos, el jugador pierde* n *pesos. a) Encuentre la esperanza de la ganancia del jugador. b) ¿Cuál es el más pequeño valor de* n *con el cual la esperanza de la ganancia es positiva?*

EJERCICIO 9.7. *Una urna contiene* N *tarjetas numeradas del* 1 *al* N. *Un experimento aleatorio consiste en seleccionar* n *tarjetas al azar y sin reemplazo. Encuentre el valor esperado del número más grande en la muestra.*

EJERCICIO 9.8. *Una urna contiene* N *tarjetas numeradas del* 1 *al* N. *Un experimento aleatorio consiste en seleccionar* n *tarjetas al azar y con reemplazo. Encuentre el valor esperado del número más grande en la muestra.*

EJERCICIO 9.9. *Sea* X *una variable aleatoria distribuida geometricamente con parámetro* p *y sea* M *un entero positivo. Encuentre la esperanza de* $Y = \text{mín}(X, M)$ *y* $Z = \text{máx}(X, M)$.

EJERCICIO 9.10. *Se escoge al azar un número del conjunto* $\{1, 2, \ldots, 10\}$. *Una persona* A *trata entonces de determinar el valor del número seleccionado mediante preguntas cuyas únicas respuestas son sí o no. Encuentre la esperanza del número de preguntas que se tienen que hacer hasta encontrar el valor del número en cada uno de los casos siguientes:*

a) La pregunta número i *es ¿se seleccionó el número* i?

b) Se pregunta de tal manera que la respuesta elimina aproximadamente la mitad de las posibles soluciones.

EJERCICIO 9.11. *Una caja contiene* 5 *componentes eléctricos, de los cuales* 2 *están defectuosos. Se checa cada uno de los componentes en un orden aleatorio hasta que se encuentren los dos defectuosos. ¿Cuál es la esperanza del número de chequeos que se realizan?*

EJERCICIO 9.12. *Se van colocando al azar bolas, una a una, en cualquiera de* N *cajas hasta que se ocupe la primera caja. ¿Cuál es el número esperado de bolas que se colocan en las cajas?*

EJERCICIO 9.13. *Una persona dispone de* n *llaves, una de las cuales es la única que abre una puerta. La persona va probando, una a una, cada llave, hasta encontrar la que abre la puerta. Encuentre el número esperado de intentos que realiza en cada uno de los casos siguientes: a) en cada intento, elige una llave al azar entre las* n, *b) en cada intento, elige una llave al azar entre las que aún no ha probado.*

EJERCICIO 9.14. *Se realizan sucesivamente ensayos de Bernoulli independientes, en cada uno de los cuales la probabilidad de éxito es* p, *hasta obtener* 2 *veces consecutivas éxito o dos veces consecutivas fracaso ¿Cuál es el número esperado de ensayos que se realizan?*

Ejercicio 9.15. *Una urna contiene N bolas numeradas del 1 al N. Se van sacando bolas al azar de la urna, una a una y con reemplazo, hasta obtener dos veces consecutivas la misma bola. ¿Cuál es el número esperado de elecciones que se realizan?*

Ejercicio 9.16. *Se quiere seleccionar a una persona de entre n. Para ello, cada una lanza una moneda y se conviene en que si alguna obtiene un resultado distinto al de las otras, entonces esa persona es la seleccionada; de otra manera se vuelven a lanzar las monedas bajo las mismas condiciones. Encuentre la esperanza del número de intentos que se realizan hasta que la persona sea seleccionada.*

Ejercicio 9.17. *Una pareja de recién casados planea tener tantos hijos como sea necesario hasta tener una niña. Suponiendo que la probabilidad de que un hijo que procree la pareja sea niña es igual a $\frac{1}{2}$, ¿cuál es el valor esperado del número de hijos que tendrá la pareja?*

Ejercicio 9.18. *Una pareja de recién casados planea tener tantos hijos como sea necesario hasta tener por lo menos una niña y por lo menos un niño, deteniendo la procreación una vez conseguido su objetivo. Suponiendo que la probabilidad de que un hijo que procree la pareja sea niña es igual a la probabilidad de que sea niño, ¿cuál es la esperanza del número de hijos que tendrá la pareja?*

***Sugerencia**: Defina Y como el número de hijos que tiene la pareja y el evento A: 'el primer hijo de la pareja es una niña'. Considere entonces la distribución de Y dado que A ocurre y la distribución de Y dado que A no ocurre.*

Ejercicio 9.19. *Cuando una determinada máquina no está bien ajustada, la probabilidad de que produzca un artículo defectuoso es igual a 0.2. Cada día la máquina se pone a funcionar y se detiene para su ajuste en el momento en que ya ha producido 3 artículos defectuosos. ¿Cuál es el número esperado de artículos que produce una máquina desajustada antes de ser detenida para su ajuste?*

Ejercicio 9.20. *Un lote contiene* $100,000$ *artículos, de los cuales* $10,000$ *están defectuosos. Se van seleccionando artículos de la población, uno a uno y al azar, hasta obtener* 20 *no defectuosos. Estime el número esperado de artículos que se inspeccionan.*

Ejercicio 9.21 (**Problema de los 3 jugadores**). *Tres jugadores, P, Q y R, juegan partidas por parejas en cada una de las cuales la probabilidad que cada jugador tiene de ganar es $\frac{1}{2}$; quien gane una partida juega con el otro jugador, hasta que uno de los jugadores gane dos partidas consecutivas, ganando entonces el juego. Encuentre el valor esperado del número de partidas que se juegan hasta que uno de los jugadores gana el juego.*

Ejercicio 9.22. *Se tienen dos muebles, cada uno con* 20 *cajones. Se elige un mueble al azar y luego un cajón al azar, de ese mueble, colocando ahí un documento.*

Una persona, que está buscando el documento, elige uno de los muebles al azar y, en ese mueble, va abriendo los cajones al azar, uno a uno, buscando ahí el documento. Si después de revisar 10 *cajones aún no se encuentra el documento, ¿qué estrategia hace mínimo el número esperado de cajones que se abrirán hasta encontrar el documento: a) continuar buscándolo en el mismo mueble y, en caso de no encontrarlo ahí, continuar con el otro o b) continuar buscándolo en el otro mueble y, en caso de no encontrarlo ahí, continuar con los diez cajones que restan del primero?*

Ejercicio 9.23. *Se tienen dos muebles, cada uno con* 20 *cajones. Se elige uno de los muebles de tal manera que la probabilidad de que sea seleccionado el primero es igual a p, en donde $p > \frac{1}{2}$; en seguida, se elige un cajón al azar, de ese mueble, colocando ahí un documento. Una persona, que está buscando el documento, elige el primer mueble y, en ese mueble, va abriendo los cajones al azar, uno a uno, buscando ahí el documento. Si después de revisar* 10 *cajones aún no se encuentra el documento, ¿qué estrategia hace mínimo el número esperado de cajones que se abrirán hasta encontrar el documento: a) continuar buscándolo en el mismo mueble y, en caso de no encontrarlo ahí, continuar con el otro o b) continuar buscándolo en el otro mueble y, en caso de no encontrarlo ahí, continuar con los diez cajones que restan del primero?*

Ejercicio 9.24. *El tiempo de vida, en días, de cierto componente, tiene distribución exponencial de parámetro $\lambda = 0.1$. Un usuario adquiere el componente bajo la condición de que pagará* 20 *pesos por cada día de operación (o bien la parte proporcional en caso de no completarse un día) después de la primera semana de uso. a) ¿Cuál es la probabilidad de que el productor reciba más de* 100 *pesos por un componente? b) ¿Cuál es la esperanza de lo que recibe el productor por un componente?*

Ejercicio 9.25. *El tiempo, en horas, que le toma a un técnico reparar una máquina tiene distribución exponencial. Si en promedio ese técnico repara* 2 *máquinas cada hora, ¿cuál es la probabilidad de que pueda reparar* 3 *máquinas en menos de* 2 *horas?*

Ejercicio 9.26. *Cada foco producido en una cierta fábrica tiene un tiempo aleatorio de vida T, con función de densidad dada por:*

$$f(t) = \begin{cases} \lambda^2 t e^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Se sabe además que, en promedio, el tiempo de vida de cada foco es de $2,000$ *horas. Estime la probabilidad de que entre* $1,000$ *focos, seleccionados al azar de la producción de esa fábrica, haya más de* 210 *con un tiempo de vida superior a* 3000 *horas.*

Ejercicio 9.27. *Sea X una variable aleatoria con distribución binomial de parámetros n y p. Encuentre la esperanza de $(1 + X)^{-1}$.*

EJERCICIO 9.28. *Sea F la función de distribución de una variable aleatoria con distribución normal estándar y sea X una variable aleatoria con distribución exponencial de parámetro $\lambda = \sqrt{2}$. Exprese $E(e^{-X^2})$ en términos de F y encuentre su valor.*

EJERCICIO 9.29. *Sea X una variable aleatoria con distribución normal estándar. Encuentre $E\left[e^{3X-X^2}\right]$.*

EJERCICIO 9.30. *La demanda de un cierto producto, que se vende según su peso, está dada por una variable aleatoria continua X con función de distribución F_X (no necesariamente absolutamente continua). La venta de x kilos del producto produce una ganancia de ax pesos, mientras que x kilos no vendidos producen una pérdida de bx pesos, en donde x es cualquier número real no negativo. Demuestre que el valor esperado de la ganancia se maximiza comprando, para su venta, x_0 kilos del producto, en donde x_0 es tal que $F_X(x_0) = \frac{a}{a+b}$.*

EJERCICIO 9.31. *Sea Z una variable aleatoria con distribución normal estándar, $y \in \mathbb{R}$ y definamos la variable aleatoria X mediante la siguiente relación:*

$$X = \begin{cases} Z & \text{si } Z > y \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre $E[X]$.

EJERCICIO 9.32. *Un experimento aleatorio consiste en lanzar 10 dados. Sea X la suma de los números que se obtienen. Encuentre $E[X]$.*

EJERCICIO 9.33. *Cada persona de un grupo de N lanza su sombrero al aire de tal manera que cada persona recoge uno de ellos al azar. ¿Cuál es la esperanza del número de personas que obtienen el sombrero que llevaban originalmente?*

EJERCICIO 9.34. *Una caja contiene n lámparas, de las cuales r funcionan correctamente. Se checa cada una de las lámparas en un orden aleatorio hasta que se encuentre una que funcione correctamente. ¿Cuál es la esperanza del número de chequeos que se realizan?*

EJERCICIO 9.35. *Una caja contiene n componentes eléctricos, de los cuales r están defectuosos. Se checa cada uno de los componentes en un orden aleatorio hasta que se encuentran s defectuosos. ¿Cuál es la esperanza del número de chequeos que se realizan?*

EJERCICIO 9.36. *Para determinar si alguna persona tiene una cierta enfermedad es necesario realizarle un análisis de sangre. Se quiere examinar a $N = n * m$ personas y, para reducir el número de pruebas, se decide formar grupos de m personas cada uno. Las muestras de sangre de las m personas en cada grupo se mezclan y se analiza la mezcla. Si la prueba resulta negativa, se concluye que ninguna persona en el grupo tiene la enfermedad y no es necesario realizar ningún otro análisis a las personas de* *ese grupo, pero si resulta positiva, se realiza el examen a cada una de las personas del grupo. Suponiendo que cada una de las N personas tiene la enfermedad con probabilidad $p = 0.01$, encuentre la esperanza del número de análisis que se realizan, siguiendo el método descrito, para determinar si cada una de las N personas tiene la enfermedad.*

EJERCICIO 9.37. *Supongamos que elegir una persona al azar y anotar el día de su cumpleaños es equivalente a elegir al azar un día de entre los 365 del año. Encuentre el número esperado de días de nacimiento distintos en un grupo de N personas elegidas al azar.*

EJERCICIO 9.38 (**Problema del colector de cupones**). *Una urna contiene N bolas numeradas del 1 al N. Se van sacando bolas de la urna, una a una y con reemplazo, hasta que se ha seleccionado por lo menos una bola de cada número. Encuentre el tamaño esperado de la muestra que se obtiene.*

EJERCICIO 9.39. *Sea X una variable aleatoria cualquiera, demuestre que a) si $F_X(m) = \frac{1}{2}$, entonces m es mediana de X, b) Si X es una variable aleatoria continua, entonces m es mediana de X si y sólo si $F_X(m) = \frac{1}{2}$, c) si no existe algún punto $x \in \mathbb{R}$ tal que $F_X(x) = \frac{1}{2}$, entonces $m = \text{ínf}\left\{x \in \mathbb{R} : F_X(x) > \frac{1}{2}\right\} = \sup\left\{x \in \mathbb{R} : F_X(x) < \frac{1}{2}\right\}$ es la única mediana de X y d) si existe algún punto $x \in \mathbb{R}$ tal que $F_X(x) = \frac{1}{2}$ entonces un número real m es mediana de X si y sólo si $m \in [a, b]$, en donde $a = \text{ínf}\left\{x \in \mathbb{R} : F_X(x) = \frac{1}{2}\right\}$ y $b = \sup\left\{x \in \mathbb{R} : F_X(x) = \frac{1}{2}\right\}$.*

EJERCICIO 9.40. *Encuentre una mediana de X si su distribución es a) uniforme en el intervalo (a, b), b) normal con parámetros μ y σ^2 y c) exponencial con parámetro λ.*

EJERCICIO 9.41. *Sea X una variable aleatoria de esperanza finita. Demuestre que $m \in \mathbb{R}$ es mediana de X si y sólo si la función $G(x) = E\left[|X - x|\right]$ alcanza su valor mínimo en $x = m$.*

EJERCICIO 9.42. *Se desea construir una estación de bomberos en algún punto de un camino de longitud infinita. Supongamos que el punto sobre el camino en el cual se presenta un incendio es aleatorio, de tal manera que la distancia, X, desde el origen hasta dicho punto, es una variable aleatoria con distribución exponencial de parámetro λ. ¿En dónde deberá construirse la estación de tal manera que la esperanza de la distancia de la estación a un incendio sobre el camino sea mínima?*

EJERCICIO 9.43. *Sea X una variable aleatoria distribuida uniformemente en el conjunto $\{0, 1, \dots N\}$. Encuentre la esperanza y la varianza de X.*

EJERCICIO 9.44. *Sean $a < b$ dos números reales y X una variable aleatoria distribuidada uniformemente en el conjunto $\left\{x_k = a + \frac{k}{n}(b - a) \mid k \in \{1, \dots, n\}\right\}$. Encuentre la esperanza y la varianza de X.*

Ejercicio 9.45. *Sea X una variable aleatoria con distribución uniforme en el intervalo $(-2,1)$. Encuentre la esperanza y la varianza de $Y = |X+1|$.*

Ejercicio 9.46. *Sea X una variable aleatoria distribuida uniformemente en el intervalo $(-3,2)$. Encuentre la esperanza y la varianza de $Y = |X^2 - 1|$.*

Ejercicio 9.47. *Se toma una muestra con reemplazo de tamaño 4 de una urna que contiene 30 bolas numeradas del 1 al 30. Sea X la suma de los números que se obtienen. Encuentre la varianza de X.*

Ejercicio 9.48. *Supongamos que X y Y son dos variables aleatorias independientes tales que $E(X^4) = 2$, $E(Y^2) = 1$, $E(X^2) = 1$ y $E(Y) = 0$. Encuentre $Var(X^2Y)$.*

Ejercicio 9.49. *Sean X y Y dos variables aleatorias independientes de varianza finita. Exprese la varianza de $2X + 3Y$ en términos de las varianzas de X y Y.*

Ejercicio 9.50. *En un grupo de 20 personas, de las cuales 10 son hombres y 10 son mujeres, se forman 10 parejas al azar. Encuentre la esperanza y la varianza del número de parejas integradas por un hombre y una mujer, entre las 10 que se forman al azar.*

Ejercicio 9.51. *Sea X una variable aleatoria con distribución hipergeométrica de parámetros r, s y n. Encuentre la varianza de X.*

Ejercicio 9.52. *Sea X una variable aleatoria con distribución gama de parámetros α y λ. Encuentre $E\left[X^n\right]$ para cualquier $n \in \mathbb{N}$.*

Ejercicio 9.53. *Sea X una variable aleatoria y $n \in \mathbb{N}$. Si X tiene esperanza finita y $|X - E(X)|^n$ también tiene esperanza finita, se dice que X tiene momento central de orden n finito y a la cantidad $E\left[(X - E(X))^n\right]$ se le llama el momento central de orden n de X. a) Demuestre que una variable aleatoria X tiene momento central de orden n finito si y sólo si X tiene momento de orden n finito. b) Para cualquier $n \in \mathbb{N}$, encuentre el momento central de orden n de una variable aleatoria con distribución normal de parámetros μ y σ^2.*

Ejercicio 9.54. *Sea X una variable aleatoria con función de densidad tipo Poisson de parámetro λ. Utilice la desigualdad de Chebyshev para verificar las siguientes desigualdades:*

a) $P(X \leq \frac{\lambda}{2}) \leq \frac{4}{\lambda}$

b) $P(X \geq 2\lambda) \leq \frac{1}{\lambda}$

Ejercicio 9.55. *Supongamos que, para decidir si una moneda está balanceada, la lanzamos 1,000 veces, y rechazamos la moneda como balanceada si el número de soles o el número de águilas que se obtienen es mayor o igual a 550. Utilice la desigualdad de Chebyshev para determinar el máximo porcentaje de monedas balanceadas que se* *rechazan con este procedimiento. Compare el resultado con el que se obtiene aplicando el teorema de de Moivre Laplace.*

Ejercicio 9.56. *Para estimar la proporción de cierta clase de peces, en una determinada población, se toma una muestra con reemplazo de tamaño* 100 *y la proporción de peces que se obtiene en la muestra se toma como la proporción de peces en la población. Calcule la probabilidad de cometer un error mayor que* 0.05 *con esta estimación.*

Ejercicio 9.57. *De acuerdo con su experiencia, un profesor sabe que la calificación de un estudiante en su examen final es una variable aleatoria* X *con esperanza* 75*. a) Obtenga una cota superior para la probabilidad de que la calificación de un estudiante exceda* 85*. Suponga además que el profesor sabe que la varianza de* X *es igual a* 25*. b) ¿Qué se puede decir acerca de la probabilidad de que un estudiante obtenga una calificación entre* 65 *y* 85*? c) ¿Cuántos estudiantes tendrían que presentar el examen de tal manera que, con una probabilidad de por lo menos* 0.9*, el promedio de calificaciones distará de* 75 *en menos que* 5*?*

Ejercicio 9.58. *Sea* X *una variable aleatoria con distribución binomial de parámetros* $n = 30$ *y* $p = \frac{1}{2}$*. a) Encuentre el valor exacto de* $P\left[12 \leq X \leq 18\right]$*. b) Utilice la desigualdad de Chebyshev para estimar* $P\left[12 \leq X \leq 18\right]$*.*

Ejercicio 9.59. *a) Utilice el teorema de de Moivre-Laplace para estimar el más pequeño valor de* n *con la propiedad de que al lanzar* n *veces una moneda, la probabilidad de que el porcentaje de las veces en que se obtiene sol esté comprendido entre* 49% *y* 51% *sea mayor o igual a* 0.95*. b) Resuelva el mismo problema utilizando la desigualdad de Chebyshev.*

Ejercicio 9.60. *Sea* X *una variable aleatoria discreta con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} & \text{si } x \in \{1, 2, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde λ *es una constante positiva. Encuentre la función generadora de probabilidades,* $\Phi_X(t) = E\left[t^X\right]$*, de* X*.*

Ejercicio 9.61. *Sea* X *una variable aleatoria discreta con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{1}{2^x} & \text{si } x \in \{1, \ldots, N\} \\ \frac{1}{2^N} & \text{si } x = N + 1 \\ 0 & \text{en otro caso} \end{cases}$$

en donde N *es un entero positivo. Encuentre la función generadora de probabilidades de* X*.*

Ejercicio 9.62. *Sea X una variable aleatoria distribuida uniformemente en el conjunto $\{0, 1, \ldots N\}$. Encuentre la función generadora de probabilidades de X y utilícela para calcular su esperanza.*

Ejercicio 9.63. *Sea X una variable aleatoria con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{2N^{x-3}}{(N+2)^{x-2}} & \text{si } x \in \{3, 4, \ldots\} \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre la función generadora de probabilidades de X y utilícela para calcular su esperanza.

Ejercicio 9.64. *Sea X una variable aleatoria distribuida binomialmente con parámetros n y p. Utilice la función generadora de probabilidades de X para calcular su esperanza y su varianza.*

Ejercicio 9.65. *Sean X y Y dos variables aleatorias independientes cuyas funciones generadoras de probabilidades están dadas por $\Phi_X(t) = e^{t^2-1}$ y $\Phi_Y(t) = \frac{1}{4}(t^3 + 2t^2 + 1)$, respectivamente. Encuentre $E[XY(X+1)]$.*

Ejercicio 9.66. *Sean X y Y dos variables aleatorias independientes cuyas funciones generadoras de probabilidades están dadas por $\Phi_X(t) = \frac{2t^4+t^3+3t+2}{8}$ y $\Phi_X(t) = \frac{4t^3+2t^2+3t}{9}$, respectivamente. Encuentre $E[XY^2(2X+3)]$.*

Ejercicio 9.67. *Sean X y Y dos variables aleatorias independientes cuyas funciones generadoras de probabilidades están dadas por $\Phi_X(t) = \frac{3t^4+t^3+5t^2+1}{10}$ y $\Phi_Y(t) = \frac{4t^3+2t^2+3t}{9}$, respectivamente. Encuentre la función de densidad de $X+Y$.*

Ejercicio 9.68. *Demuestre los siguientes resultados:*

a) Si X y Y son dos variables aleatorias independientes, ambas con distribución binomial de parámetros n_1, p y n_2, p, respectivamente, entonces $X+Y$ tiene distribución binomial con parámetros $n_1 + n_2, p$.

b) Si X y Y son dos variables aleatorias independientes, ambas con distribución Poisson de parámetros λ_1 y λ_2, respectivamente, entonces $X+Y$ tiene distribución Poisson con parámetro $\lambda_1 + \lambda_2$.

Ejercicio 9.69. *Sea X una variable aleatoria con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{1}{2}e^{x} & \text{si } x \leq 0 \\ \frac{1}{2}e^{-x} & \text{si } x > 0 \end{cases}$$

Encuentre la función generadora de momentos de X, especificando la región en la cual está bien definida.

Ejercicio 9.70. *Demuestre que si* X *y* Y *son variables aleatorias independientes, ambas con distribución normal de parámetros* μ_1, σ_1^2 *y* μ_2, σ_2^2*, respectivamente, entonces* $X+Y$ *tiene distribución normal con parámetros* $\mu_1+\mu_2, \sigma_1^2+\sigma_2^2$.

Ejercicio 9.71. *Sean* X *y* Y *dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo* $(0,1)$. *Encuentre la función generadora de momentos de a)* $U = X+Y$ *y b)* $V = X-Y$.

Ejercicio 9.72. *Utilice la función generadora de momentos para encontrar la esperanza y la varianza de* X *en cada uno de los casos siguientes:*

a) X *tiene distribución exponencial con parámetro* λ.

b) X *tiene distribución gama con parámetros* α *y* λ.

c) X *tiene distribución normal con parámetros* μ *y* σ^2.

Ejercicio 9.73. *Sean* X *y* Y *dos variables aleatorias independientes cuyas funciones generadoras de momentos están dadas por* $M_X(t) = (1-t)^{-3}\, e^{2t}$ *y* $M_Y(t) = \frac{1}{4}\left(e^{3t}+2e^{2t}+1\right)$*, respectivamente. Encuentre* $E\left[X^3 Y (X - Y^2)\right]$.

Apéndice

Los desarrollos de Riemann sobre las series trigonométricas deben ser vistos, en cuanto a los principios fundamentales, como perfectamente de acuerdo con los métodos de exposición de Wierstrass, quien, en este aspecto, suprime completamente la intuición geométrica y opera exclusivamente con las definiciones aritméticas. Pero no me puedo imaginar que Riemann, en su interior, haya visto alguna vez la intuición geométrica, al igual que lo hacen ciertos representantes celosos de las Matemáticas hipermodernas, como algo contrario al espíritu matemático y que deba conducir necesariamente a conclusiones erróneas. Por el contrario, él debió pensar que en las dificultades que se presentan aquí, es posible encontrar un terreno de conciliación.

Félix Christian Klein

A.1. Sucesiones de números reales

DEFINICIÓN 1 (**Sucesión de números reales**). *Una sucesión de números reales es una función* $f : \mathbb{N} \mapsto \mathbb{R}$. *Si, para* $n \in \mathbb{N}$, $x_n = f(n)$, *se denota a la sucesión por* $(x_n)_{n\in\mathbb{N}}$ *o simplemente por* (x_n).

DEFINICIÓN 2 (**Sucesión creciente**). *Se dice que una sucesión* (x_n) *es creciente (resp. estrictamente creciente) si* $x_n \leq x_{n+1}$ *(resp.* $x_n < x_{n+1}$*) para cualquier* $n \in \mathbb{N}$.

DEFINICIÓN 3 (**Sucesión decreciente**). *Se dice que una sucesión* (x_n) *es decreciente (resp. estrictamente decreciente) si* $x_n \geq x_{n+1}$ *(resp.* $x_n > x_{n+1}$*) para cualquier* $n \in \mathbb{N}$.

DEFINICIÓN 4 (**Subsucesión**). *Sea (x_n) una sucesión de números reales y (r_n) una sucesión de números naturales estrictamente creciente, entonces, la sucesión (x_{r_n}) es llamada una subsucesión de (x_n).*

DEFINICIÓN 5 (**Convergencia de una sucesión**). *Se dice que una sucesión (x_n) converge si existe un número real x tal que, para cualquier $\varepsilon > 0$, existe $N \in \mathbb{N}$ tal que $|x_n - x| < \varepsilon$ para cualquier $n \geq N$. Si la sucesión no converge, se dice que diverge. Si la sucesión converge a x escribiremos $\lim_{n\rightsquigarrow\infty} x_n = x$.*

La siguiente proposición se demuestra fácilmente:

PROPOSICIÓN 6. *Sean (x_n) y (y_n) dos sucesiones convergentes y $c \in \mathbb{R}$, entonces:*

(i) $\lim_{n\rightsquigarrow\infty} cx_n = c\lim_{n\rightsquigarrow\infty} x_n$
(ii) $\lim_{n\rightsquigarrow\infty} (x_n + y_n) = \lim_{n\rightsquigarrow\infty} x_n + \lim_{n\rightsquigarrow\infty} +y_n$
(iii) $\lim_{n\rightsquigarrow\infty} x_n y_n = \lim_{n\rightsquigarrow\infty} x_n \lim_{n\rightsquigarrow\infty} y_n$
(iv) *Si* $\lim_{n\rightsquigarrow\infty} y_n \neq 0$, *entonces* $\lim_{n\rightsquigarrow\infty} \frac{1}{y_n} = \frac{1}{\lim_{n\rightsquigarrow\infty} y_n}$
(v) *Si* $x_n \leq y_n$ *a partir de alguna* n, *entonces* $\lim_{n\rightsquigarrow\infty} x_n \leq \lim_{n\rightsquigarrow\infty} y_n$

DEFINICIÓN 7 (**Divergencia de una sucesión a** ∞ o $-\infty$). *Sea (x_n) una sucesión de números reales. Diremos que la sucesión diverge a ∞ (resp. $-\infty$) si, para cualquier $M \in \mathbb{R}$, existe $N \in \mathbb{N}$ tal que $x_n \geq M$ para cualquier $n \geq N$. En ese caso, escribiremos $\lim_{n\rightsquigarrow\infty} x_n = \infty$ (resp. $\lim_{n\rightsquigarrow\infty} x_n = -\infty$).*

DEFINICIÓN 8 (**Sucesión acotada**). *Se dice que una sucesión (x_n) está acotada por abajo (resp. por arriba) si existe un número real M tal que $x_n \leq M$ (resp. $x_n \geq M$) para cualquier $n \in \mathbb{N}$. Se dice que una sucesión está acotada si lo está por arriba y por abajo.*

DEFINICIÓN 9 (**Punto límite de una sucesión**). *Sea (x_n) una sucesión de números reales. Se dice que un número real x es punto límite de (x_n) si existe una subsucesión de (x_n) que converge a x. También diremos que ∞ (resp. $-\infty$) es punto límite de (x_n) si existe una subsucesión de (x_n) que diverge a ∞ (resp. $-\infty$).*

TEOREMA 10 (**Teorema de Bolzano-Weierstrass**). *Si la sucesión (x_n) está acotada entonces existe por lo menos un número real x que es punto límite de (x_n).*

Demostración

Sea $M \in \mathbb{R}$ tal que $|x_n| \leq M$ para cualquier $n \in \mathbb{N}$.

Definamos inductivamente una sucesión de intervalos $I_n = [a_n, b_n]$ tales que:

(i) $b_n - a_n = \frac{M}{2^{n-2}}$
(ii) $I_{n+1} \subset I_n$ para cualquier $n \in \mathbb{N}$

(iii) Para cualquier $n \in \mathbb{N}$, I_n contiene una infinidad de términos de la sucesión (x_n).

Esta definición puede hacerse de la siguiente manera:

Definamos $I_1 = [-M, M]$

Habiendo definido el intervalo $I_n = [a_n, b_n]$ con las propiedades mencionadas, sea c_n el punto medio de I_n. Entonces, al menos uno de los intervalos $[a_n, c_n]$ ó $[c_n, b_n]$ contiene una infinidad de términos de la sucesión (x_n). Definamos entonces I_{n+1} cualquiera de esos dos intervalos que contenga una infinidad de términos de la sucesión (x_n).

La sucesión de intervalos I_n así construida satisface las propiedades mencionadas.

Por el teorema de los intervalos encajados, se tiene:

$\bigcap_{n=1}^{\infty} I_n \neq \emptyset$

Además, como la longitud de los intervalos I_n tiende a cero, la intersección $\bigcap_{n=1}^{\infty} I_n$ contiene un único punto x.

Tomemos en cada intervalo I_n un elemento x_{k_n} de la sucesión (x_n) de tal forma que $k_{n+1} > k_n$. Entonces, (x_{k_n}) es una subsucesión de (x_n) que converge a x.

■

Si una sucesión (x_n) no está acotada, entonces existe una subsucesión que diverge a ∞ o a $-\infty$, así que se tiene el siguiente resultado:

Corolario 11. *Toda sucesión tiene por lo menos un punto límite.*

Proposición 12. *Una sucesión (x_n) converge a $x \in \mathbb{R}$ si y sólo si x es el único punto límite de (x_n).*

Demostración

Es inmediato que si (x_n) converge a x entonces toda subsucesión de (x_n) también converge a x, así que x es el único punto límite de (x_n).

Supongamos ahora que x es el único punto límite de (x_n). Esto implica, en particular que (x_n) está acotada pues, de otra forma, ∞ o $-\infty$ sería también punto límite de (x_n).

Supongamos que (x_n) no converge a x, entonces existe $\varepsilon > 0$ tal que, para cualquier $N \in \mathbb{N}$, existe $n \geq N$ tal que $|x_n - x| \geq \varepsilon$. Se puede entonces tomar una sucesión creciente de números naturales (k_n) tal que $|x_{k_n} - x| \geq \varepsilon$ para cualquier $n \in \mathbb{N}$. Por el teorema de Bolzano-Weierstrass, existe un número real y que es punto límite de

(x_{k_n}). Se tiene $|y - x| \geq \varepsilon$, así que $y \neq x$, lo cual contradice el que x es el único punto límite de (x_n).

∎

COROLARIO 13. *Si* (x_n) *es una sucesión creciente (resp. decreciente) acotada por arriba (resp. acotada por abajo), entonces es convergente.*

DEFINICIÓN 14 (**Sucesión de Cauchy**). *Se dice que una sucesión* (x_n) *es de Cauchy si para cualquier* $\varepsilon > 0$*, existe* $N \in \mathbb{N}$ *tal que* $|x_n - x_m| < \varepsilon$ *para cualesquiera* $n, m \geq N$.

TEOREMA 15. *Una sucesión* (x_n) *es convergente si y sólo si es de Cauchy.*

Demostración

Que una sucesión convergente es de Cauchy se demuestra fácilmente.

Inversamente, supongamos que (x_n) es de Cauchy y sea $N \in \mathbb{N}$ tal que $|x_n - x_m| < 1$ para cualesquiera $n, m \geq N$. Entonces $|x_n| \leq |x_n - x_N| + |x_N| < 1 + |x_N|$ para cualquier $n \geq N$, así que $\max\{|x_1|, \dots, |x_{N-1}|, 1 + |x_N|\}$ es una cota de la sucesión (x_n).

Por el teorema de Bolzano-Weierstrass, existe un número real x y una subsucesión (x_{k_n}) de (x_n) tal que $\lim_{n \rightsquigarrow \infty} x_{k_n} = x$.

Sea $y \in \mathbb{R}$ un punto límite de (x_n) y (x_{j_n}) una subsucesión de (x_n) tal que $\lim_{n \rightsquigarrow \infty} x_{j_n} = y$.

Dada $\varepsilon > 0$ sean $N_1 \in \mathbb{N}$ tal que $|x_{k_n} - x| < \frac{\varepsilon}{3}$ para cualquier $n \geq N_1$, $N_2 \in \mathbb{N}$ tal que $|x_{j_n} - y| < \frac{\varepsilon}{3}$ para cualquier $n \geq N_2$ y $N_3 \in \mathbb{N}$ tal que $|x_n - x_m| < \frac{\varepsilon}{3}$ para cualesquiera $n, m \geq N_3$. Entonces, tomando k_n y j_n tales que $n \geq \max\{N_1, N_2\}$ y $k_n, j_n \geq N_3$, se tiene:

$$|x - y| \leq |x_{k_n} - x| + |x_{k_n} - x_{j_n}| + |x_{j_n} - y| < \varepsilon$$

Así que $|x - y| < \varepsilon$ para cualquier $\varepsilon > 0$, por consiguiente, $x = y$.

Por lo tanto, utilizando la proposición 12, se concluye que la sucesión (x_n) es convergente.

∎

Haciendo la convención $-\infty < x < \infty$ para cualquier número real x, todo subconjunto de $\mathbb{R} \cup \{-\infty, \infty\}$ tiene un supremo y un ínfimo, pudiendo ser éstos ∞ o $-\infty$. Denotaremos por $\overline{\mathbb{R}}$ al conjunto $\mathbb{R} \cup \{-\infty, \infty\}$.

DEFINICIÓN 16 (**Límites superior e inferior de una sucesión**). *Sea* (x_n) *una sucesión de números reales, entonces definimos el límite superior de* (x_n)*,* $\limsup x_n$*, y el límite inferior de* (x_n)*,* $\liminf x_n$*, de la siguiente manera:*

(i) $\liminf x_n = \inf\left\{x \in \overline{\mathbb{R}} : x \text{ es punto límite de } (x_n)\right\}$
(ii) $\limsup x_n = \sup\left\{x \in \overline{\mathbb{R}} : x \text{ es punto límite de } (x_n)\right\}$

Se tiene siempre $\liminf x_n \leq \limsup x_n$ y la proposición 12 implica inmediatamente el siguiente resultado:

PROPOSICIÓN 17. *Una sucesión* (x_n) *es convergente si y sólo si:*

$\liminf x_n = \limsup x_n \in \mathbb{R}$

PROPOSICIÓN 18 (**Caracterización de los límites superior e inferior**). *Para cualquier sucesión* (x_n) *se tiene:*

(i) $\liminf x_n$ *es punto límite de* (x_n) *y:*
$\liminf x_n = \lim_{n\rightsquigarrow\infty} (\inf_{j\geq n}\{x_j\})$
(ii) $\limsup x_n$ *es punto límite de* (x_n) *y:*
$\limsup x_n = \lim_{n\rightsquigarrow\infty} \left(\sup_{j\geq n}\{x_j\}\right)$

Demostración

La sucesión $z_n = \inf_{j\geq n}\{x_j\}$ es creciente así que el límite $z = \lim_{n\rightsquigarrow\infty} z_n$ existe (pudiendo ser ∞).

Si $z = -\infty$, entonces, dada cualquier $M \in \mathbb{R}$, existe $N \in \mathbb{N}$ tal que $z_n < M$ para cualquier $n \geq N$. Existe entonces una subsucesión (z_{k_n}) tal que $z_{k_n} < -n$ para cualquier $n \in \mathbb{N}$, es decir, $\inf_{j\geq k_n}\{x_j\} < -n$. Por lo tanto, para cada $n \in \mathbb{N}$, existe $j_n \geq k_n$ tal que $x_{j_n} < -n$. Finalmente, como la sucesión (k_n) es estrictamente creciente, existe una sucesión estrictamente creciente, (j_{i_n}) tal que $x_{j_{i_n}} < -i_n$ para cualquier $n \in \mathbb{N}$. Por lo tanto, $\lim_{n\rightsquigarrow\infty} x_{j_{i_n}} = -\infty$. Así que $-\infty$ es punto límite de la sucesión (x_n).

Si $z = \infty$, entonces, dada cualquier $M \in \mathbb{R}$, existe $N \in \mathbb{N}$ tal que $z_n > M$ para cualquier $n \geq N$, así que $x_n > M$ para cualquier $n \geq N$, es decir, $\lim_{n\rightsquigarrow\infty} x_n = \infty$.

Si $z \in \mathbb{R}$, entonces, dada cualquier $\varepsilon > 0$, existe $N \in \mathbb{N}$ tal que $z - \varepsilon < z_n \leq z$ para cualquier $n \geq N$. Existe entonces una subsucesión (z_{k_n}) tal que $z - \frac{1}{n} < z_{k_n} \leq z$ para cualquier $n \in \mathbb{N}$, es decir, $z - \frac{1}{n} < \inf_{j\geq k_n}\{x_j\} \leq z$. Por lo tanto, para cada $n \in \mathbb{N}$, existe $j_n \geq k_n$ tal que $z - \frac{1}{n} < x_{j_n} \leq z$. Finalmente, como la sucesión (k_n) es estrictamente creciente, existe una sucesión estrictamente creciente, (j_{i_n}) tal que

$z - \frac{1}{i_n} < x_{j_{i_n}} \leq z$ para cualquier $n \in \mathbb{N}$. Por lo tanto, $\text{lím}_{n \rightsquigarrow \infty} x_{j_{i_n}} = z$. Así que z es punto límite de la sucesión (x_n).

Por lo tanto, en cualquier caso, $z = \text{lím}_{n \rightsquigarrow \infty} \left(\text{ínf}_{j \geq n} \{x_j\} \right)$ es punto límite de la sucesión (x_n).

Si $z = -\infty$, entonces, $\text{lím ínf}\, x_n = -\infty$, así que $z = \text{lím ínf}\, x_n$.

Si $z = \infty$, entonces, como $\text{lím}_{n \rightsquigarrow \infty} x_n = \infty$, ∞ es el único punto límite de (x_n), así que $z = \text{lím ínf}\, x_n$.

Si $z \in \mathbb{R}$, entonces, dada cualquier $\varepsilon > 0$, existe $N \in \mathbb{N}$ tal que $z_n > z - \varepsilon$ para cualquier $n \geq N$, así que $x_n > z - \varepsilon$ para cualquier $n \geq N$. Por lo tanto, no existe alguna subsucesión de (x_n) que diverja a $-\infty$ y, si (x_{k_n}) es una subsucesión convergente de (x_n), se tiene: $\text{lím}_{n \rightsquigarrow \infty} x_{k_n} \geq z - \varepsilon$ para cualquier $\varepsilon > 0$, así que $\text{lím}_{n \rightsquigarrow \infty} x_{k_n} \geq z$. Por lo tanto, $z = \text{lím ínf}\, x_n$.

Para la otra parte, x es punto límite de (x_n) si y sólo si $-x$ es punto límite de $(-x_n)$, así que:

$\text{lím}_{n \rightsquigarrow \infty} \left(\sup_{j \geq n} \{x_j\} \right) = -\text{lím}_{n \rightsquigarrow \infty} \left(\text{ínf}_{j \geq n} \{-x_j\} \right)$

es punto límite de (x_n).

Además:

$\text{lím sup}\, x_n = -\text{lím ínf}\, (-x_n) = -\text{lím}_{n \rightsquigarrow \infty} \left(\text{ínf}_{j \geq n} \{-x_j\} \right)$

$= -\text{lím}_{n \rightsquigarrow \infty} \left(-\sup_{j \geq n} \{x_j\} \right) = \text{lím}_{n \rightsquigarrow \infty} \left(\sup_{j \geq n} \{x_j\} \right)$

■

A.2. Series de números reales

DEFINICIÓN 19 (**Serie de números reales**). *Si (x_n) es una sucesión de números reales, la serie generada por (x_n) es la sucesión (s_n) definida por $s_n = \sum_{i=1}^{n} x_i$ y se le denota por $\sum_{i=1}^{\infty} x_i$. Se dice que la serie converge si converge la sucesión (s_n). En caso contrario, se dice que la serie diverge.*

PROPOSICIÓN 20. *Si la serie $\sum_{i=1}^{\infty} x_i$ converge, entonces las sucesiones (x_n) y $(\sum_{i=n}^{\infty} x_i)$ convergen a 0.*

Demostración

Sea $s_n = \sum_{i=1}^{n} x_i$, entonces $x_n = s_n - s_{n-1}$, así que:

$\lim_{n\rightsquigarrow\infty} x_n = \lim_{n\rightsquigarrow\infty} s_n - \lim_{n\rightsquigarrow\infty} s_{n-1} = 0$

Por otra parte, para cualquier $n \in \mathbb{N}$ y $m \geq n$, se tiene:

$\sum_{i=n}^{m} x_i = \sum_{i=1}^{m} x_i - \sum_{i=1}^{n-1} x_i = s_m - s_{n-1}$

Así que la serie $\sum_{i=n}^{\infty} x_i$ converge y $\sum_{i=n}^{\infty} x_i = \lim_{m\rightsquigarrow\infty} s_m - s_{n-1}$. Por lo tanto:

$\lim_{n\rightsquigarrow\infty} \sum_{i=n}^{\infty} x_i = \lim_{n\rightsquigarrow\infty} (\lim_{m\rightsquigarrow\infty} s_m - s_{n-1})$

$= \lim_{m\rightsquigarrow\infty} s_m - \lim_{n\rightsquigarrow\infty} s_{n-1} = 0$

■

DEFINICIÓN 21 (**Convergencia absoluta de una serie**). *Se dice que la serie (s_n), generada por (x_n), converge absolutamente si la serie $\sum_{k=1}^{\infty} |x_n|$ converge.*

La convergencia absoluta de una serie implica su convergencia. Cuando la sucesión (x_n) está formada por términos no negativos, la convergencia de la serie $\sum_{k=1}^{\infty} x_n$ es equivalente a su convergencia absoluta. Además, en ese caso, la serie converge si y sólo si la sucesión $s_n = \sum_{k=1}^{n} x_n$ está acotada, en cuyo caso la serie converge a $\sup\{s_n \mid n \in \mathbb{N}\}$; en caso de que esto no ocurra, se dice que la serie diverge a ∞, lo cual se denota por $\sum_{k=1}^{\infty} x_n = \infty$.

PROPOSICIÓN 22. *Sea $\sum_{i=1}^{\infty} x_i$ una serie absolutamente convergente y $k_1, k_2, \ldots$ cualquier ordenamiento de los números naturales, entonces la serie $\sum_{i=1}^{\infty} x_{k_i}$ es absolutamente convergente y $\sum_{i=1}^{\infty} x_{k_i} = \sum_{i=1}^{\infty} x_i$.*

Demostración

Para cada $n \in \mathbb{N}$, sea $j_n = \max\{k_1, \ldots, k_n\}$, entonces:

$\sum_{i=1}^{n} |x_{k_i}| \leq \sum_{i=1}^{j_n} |x_i| \leq \sum_{i=1}^{\infty} |x_i| < \infty$

Así que la serie $\sum_{i=1}^{\infty} x_{k_i}$ es absolutamente convergente.

Sea $S = \sum_{i=1}^{\infty} x_i$ y $S' = \sum_{i=1}^{\infty} x_{k_i}$.

Dada $\varepsilon > 0$, sea $N_1 \in \mathbb{N}$ tal que $\sum_{i=n}^{\infty} |x_i| < \frac{\varepsilon}{3}$ para cualquier $n \geq N_1$ y $N_2 \in \mathbb{N}$ tal que $\sum_{i=n}^{\infty} |x_{k_i}| < \frac{\varepsilon}{3}$ para cualquier $n \geq N_2$.

Sea $N \geq N_2$ tal que:

$\{k_1, \ldots, k_{N-1}\} \supset \{1, \ldots, N_1 - 1\}$

Entonces:

$|S-S'| = \left|S - \sum_{i=1}^{N_1-1} x_i\right| + \left|\sum_{i=1}^{N-1} x_{k_i} - \sum_{i=1}^{N_1-1} x_i\right| + \left|S' - \sum_{i=1}^{N-1} x_{k_i}\right|$

$\leq \left|\sum_{i=N_1}^{\infty} x_i\right| + \sum_{i=N_1}^{\infty} |x_i| + \left|\sum_{i=N}^{\infty} x_{k_i}\right|$

$\leq 2\sum_{i=N_1}^{\infty} |x_i| + \sum_{i=N}^{\infty} |x_{k_i}| < \varepsilon$

Así que $|S-S'| < \varepsilon$ para cualquier $\varepsilon > 0$, por lo tanto, $S = S'$.

■

Proposición 23. *Para cada pareja $j,k \in \mathbb{N}$ sea a_{jk} un número real no negativo y b_1, b_2, ... cualquier ordenamiento de los elementos a_{jk}. Entonces:*

$$\sum_{i=1}^{\infty} b_i = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk} = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} a_{jk}$$

Demostración

Si la serie $\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk}$ converge a L, para cualquier pareja $n,m \in \mathbb{N}$, se tiene $\sum_{j=1}^{n}\sum_{k=1}^{m} a_{jk} \leq L$. De manera que, dado $N \in \mathbb{N}$, podemos tomar $n,m \in \mathbb{N}$ tales que $\{b_1,\ldots,b_N\} \subset \{a_{jm} \mid j \leq n,\ k \leq m\}$ y entonces $\sum_{i=1}^{N} b_i \leq \sum_{j=1}^{n}\sum_{k=1}^{m} a_{jk} \leq L$. Así que la serie $\sum_{i=1}^{\infty} b_i$ converge y $\sum_{i=1}^{\infty} b_i \leq \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk}$.

Supongamos ahora que la serie $\sum_{i=1}^{\infty} b_i$ converge a L, entonces, para cualquier pareja $n,m \in N$, se tiene $\sum_{j=1}^{n}\sum_{k=1}^{m} a_{jk} \leq \sum_{i=1}^{\infty} b_i = L$, así que $\sum_{j=1}^{n}\sum_{k=1}^{\infty} a_{jk} \leq L$ para cualquier $n \in N$ y, por lo tanto, $\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk} \leq L$. Así que la serie $\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk}$ converge y $\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk} \leq \sum_{i=1}^{\infty} b_i$.

Se concluye entonces que la serie $\sum_{i=1}^{\infty} b_i$ converge si y sólo si la serie $\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk}$ converge y, en ese caso, $\sum_{i=1}^{\infty} b_i = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk}$.

De la misma manera, la serie $\sum_{i=1}^{\infty} b_i$ converge si y sólo si la serie $\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} a_{jk}$ converge y, en ese caso, $\sum_{i=1}^{\infty} b_i = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} a_{jk}$.

■

Definición 24 (**Serie de potencias**). *Si $(a_n)_{n\in\mathbb{N}}$ es una sucesión de números reales, la serie $\sum_{k=1}^{\infty} a_n x^n$, definida para $x \in \mathbb{R}$, es llamada una serie de potencias.*

Las demostraciones de los siguientes dos resultados pueden consultarse en el libro de Robert G. Bartle, Introducción al Análisis Matemático, p. 351-358, Limusa, 1990.

Teorema 25. *Sea (a_n) una sucesión de números reales y $R > 0$. Supongamos que la serie de potencias $\sum_{n=1}^{\infty} a_n x^n$ converge absolutamente para $x \in (-R,R)$, entonces la serie $\sum_{n=1}^{\infty} n a_n x^{n-1}$ también converge absolutamente para $x \in (-R,R)$. Además,* *si $f : (-R, R) \mapsto \mathbb{R}$ está definida por $f(x) = \sum_{n=1}^{\infty} a_n x^n$, entonces f es continua y derivable en el intervalo $(-R, R)$ y $f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$.*

TEOREMA 26. *Sea (a_n) una sucesión de números reales. Supongamos que la serie de potencias $\sum_{n=1}^{\infty} a_n x^n$ converge absolutamente para $x \in [0,1]$ y sea $f : [0,1] \mapsto \mathbb{R}$ definida por $f(x) = \sum_{n=1}^{\infty} a_n x^n$, entonces $\sum_{n=1}^{\infty} a_n = \lim_{x \rightsquigarrow 1-} f(x)$.*

A.3. La integral de Riemann

DEFINICIÓN 27 (**Partición de un intervalo**). *Una partición P del intervalo $[a, b]$ es un conjunto finito de números reales $\{x_0, \ldots, x_n\}$ tales que $a = x_0 < x_1 < \cdots < x_n = b$.*

DEFINICIÓN 28 (**Refinamiento de una partición**). *Se dice que una partición P es un refinamiento de la partición Q si $P \supset Q$.*

DEFINICIÓN 29 (**Suma de Riemann**). *Dada una función acotada $f : [a, b] \mapsto \mathbb{R}$ y una partición P del intervalo $[a, b]$, una suma de Riemann de f con respecto a P es un número $S(P; f)$ definido por:*

$$S(P; f) = \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$

en donde $\xi_i \in [x_{i-1}, x_i]$.

DEFINICIÓN 30 (**Función Riemann integrable**). *Se dice que una función acotada $f : [a, b] \mapsto \mathbb{R}$ es Riemann integrable en el intervalo $[a, b]$ si existe un número real I tal que, para cualquier $\varepsilon > 0$, existe una partición P_ε del intervalo $[a, b]$ tal que si P es un refinamiento de P_ε entonces:*

$$|S(P; f) - I| < \varepsilon$$

para cualquier suma de Riemann, $S(P; f)$, de f con respecto a P.

Al numero real I de esta definición se le llama la integral de Riemann de f en el intervalo $[a, b]$ y se le denota por $\int_a^b f(x)dx$.

TEOREMA 31 (**Criterio de Cauchy**). *Sea $f : [a, b] \mapsto \mathbb{R}$ una función acotada, entonces f es Riemann integrable en el intervalo $[a, b]$ si y sólo si, para cualquier $\varepsilon > 0$, existe una partición P_ε del intervalo $[a, b]$ tal que si P y P' son refinamientos de P_ε entonces:*

$$|S(P; f) - S(P; f)| < \varepsilon$$

para cualquier suma de Riemann, $S(P;f)$, de f con respecto a P y cualquier suma de Riemann, $S(P';f)$, de f con respecto a P'.

Demostración

Si f es Riemann integrable en el intervalo $[a,b]$, para cualquier $\varepsilon > 0$, existe una partición P_ε del intervalo $[a,b]$ tal que si P es un refinamiento de P_ε entonces:

$\left|S(P;f) - \int_a^b f(x)dx\right| < \frac{\varepsilon}{2}$

para cualquier suma de Riemann, $S(P;f)$, de f con respecto a P.

Si P' es otro refinamiento de P_ε, se tiene:

$|S(P;f) - S(P;f)| \leq \left|S(P;f) - \int_a^b f(x)dx\right| + \left|S(P;f) - \int_a^b f(x)dx\right| < \varepsilon$

para cualquier suma de Riemann, $S(P;f)$, de f con respecto a P y cualquier suma de Riemann, $S(P';f)$, de f con respecto a P'.

Para el inverso, definamos inductivamente una sucesión de particiones $\{Q_n\}$ tal que $Q_0 = \{a,b\}$ y, para $n \geq 1$, $Q_n \supset Q_{n-1}$ y si $P, Q \supset Q_n$ entonces $|S(P,f,g) - S(Q,f,g)| < \frac{1}{n}$ para cualquier par de sumas de Riemann $S(P;f)$ y $S(Q;f)$.

Para cada n consideremos entonces cualquier suma de Riemann $S(Q_n;f)$. La sucesión $\{S(Q_n;f)\}$ es de Cauchy; por lo tanto, converge.

Sea $I = \text{lím}_{n\rightsquigarrow\infty} S(Q_n;f)$ y, dada $\varepsilon > 0$, sea N tal que $\frac{1}{N} < \frac{\varepsilon}{2}$ y $|S(Q_n;f) - I| < \frac{\varepsilon}{2}$ para toda $n \geq N$.

Si P es un refinamiento de Q_N, se tiene $|S(P;f) - S(Q_N;f)| < \frac{1}{N}$, por lo tanto:

$|S(P;f) - I| \leq |S(P;f) - S(Q_N;f)| + |S(Q_N;f) - I| < \frac{1}{N} + \frac{\varepsilon}{2} < \varepsilon$

Se concluye entonces que f es Riemann integrable en el intervalo $[a,b]$.

■

DEFINICIÓN 32. *Sea $f : [a,b] \mapsto \mathbb{R}$ una función acotada e I un intervalo cerrado contenido en $[a,b]$. La diferencia:*

$$\sup\{f(x) : x \in I\} - \text{ínf}\{f(x) : x \in I\}$$

es llamada la oscilación de f en el intervalo I y será denotada por O_I.

PROPOSICIÓN 33. *Sea $f : [a,b] \mapsto \mathbb{R}$ una función acotada y $x \in (a,b)$. Sea (I_n) una sucesión de intervalos cerrados encajados, contenidos en $[a,b]$, que contengan a*

x como punto interior y tales que $\bigcap_{n=1}^{\infty} I_n = \{x\}$. Entonces el $\text{lím}_{n\rightsquigarrow\infty} O_{I_n}$ *existe y es independiente de la sucesión particular de intervalos encajados con las propiedades mencionadas.*

Demostración

Como los intervalos están encajados, la sucesión O_{I_n} es monótona decreciente y está acotada, por abajo, por 0; por lo tanto, es convergente.

Sea (J_n) cualquier sucesión de intervalos cerrados encajados, contenidos en $[a,b]$, que contengan a x como punto interior y tales que $\bigcap_{n=1}^{\infty} J_n = \{x\}$, y denotemos por O_{J_n} a la oscilación de f en el intervalo J_n.

Si $J_n = [c_n, d_n]$, entonces, como $J_n \supset J_{n+1}$, la sucesión (c_n) es monótona creciente y la sucesión (d_n) es monótona decreciente. Ambas sucesiones están contenidas en el intervalo $[a,b]$, así que convergen.

Sea $c = \text{lím}_{n\rightsquigarrow\infty} c_n$ y $d = \text{lím}_{n\rightsquigarrow\infty} d_n$. Entonces $\bigcap_{n=1}^{\infty} J_n = [c,d]$ y $\text{lím}_{n\rightsquigarrow\infty} (d_n - c_n) = d - c$. Por lo tanto, $c = d = x$ y la sucesión $(d_n - c_n)$ decrece a 0.

Como cada intervalo J_n contiene a x como punto interior, se tiene $d_n - x > 0$ y $x - c_n > 0$ para cualquier $n \in \mathbb{N}$.

Dado cualquier intervalo $I_k = [a_k, b_k]$, se tiene $d_k - x > 0$ y $x - c_k > 0$, de manera que, como $\text{lím}_{n\rightsquigarrow\infty} (d_n - c_n) = 0$, existe una N tal que $d_n - x < d_k - x$ y $x - c_n < x - c_k$ para cualquier $n \geq N$. Por lo tanto, $J_n \subset I_k$ para cualquier $n \geq \mathbb{N}$, lo cual implica que $O_{J_n} \leq O_{I_k}$ para cualquier $n \geq \mathbb{N}$. Por lo tanto $\text{lím}_{n\rightsquigarrow\infty} O_{J_n} \leq O_{I_k}$ para cualquier k, y entonces $\text{lím}_{n\rightsquigarrow\infty} O_{J_n} \leq \text{lím}_{n\rightsquigarrow\infty} O_{I_n}$.

De la misma forma, se tiene $\text{lím}_{n\rightsquigarrow\infty} O_{I_n} \leq \text{lím}_{n\rightsquigarrow\infty} O_{J_n}$, así que $\text{lím}_{n\rightsquigarrow\infty} O_{J_n} = \text{lím}_{n\rightsquigarrow\infty} O_{I_n}$.

■

De la misma manera, se demuestran los siguientes dos resultados:

PROPOSICIÓN 34. *Sea $f : [a,b] \mapsto \mathbb{R}$ una función acotada e (I_n) una sucesión de intervalos encajados de la forma $[a, b_n]$, con $b_n \leq b$, y tales que $\bigcap_{n=1}^{\infty} I_n = \{a\}$. Entonces el* $\text{lím}_{n\rightsquigarrow\infty} O_{I_n}$ *existe y es independiente de la sucesión particular de intervalos encajados con las propiedades mencionadas.*

PROPOSICIÓN 35. *Sea $f : [a,b] \mapsto \mathbb{R}$ una función acotada e (I_n) una sucesión de intervalos encajados de la forma $[a_n, b]$, con $a \leq a_n$, y tales que $\bigcap_{n=1}^{\infty} I_n = \{b\}$. Entonces el* $\text{lím}_{n\rightsquigarrow\infty} O_{I_n}$ *existe y es independiente de la sucesión particular de intervalos encajados con las propiedades mencionadas.*

DEFINICIÓN 36. *Sea $f : [a,b] \mapsto \mathbb{R}$ una función acotada.*

(i) *Si $x \in (a,b)$, se define la oscilación de la función f en el punto x como el límite* $\lim_{n \rightsquigarrow \infty} O_{I_n}$*, en donde (I_n) es una sucesión de intervalos cerrados encajados, contenidos en $[a,b]$, que contienen a x como punto interior y tales que $\bigcap_{n=1}^{\infty} I_n = \{x\}$.*
(ii) *Se define la oscilación de la función f en el punto a como el límite* $\lim_{n \rightsquigarrow \infty} O_{I_n}$*, en donde (I_n) es una sucesión de intervalos encajados de la forma $[a,b_n]$, con $a < b_n \leq b$, y tales que $\bigcap_{n=1}^{\infty} I_n = \{a\}$.*
(iii) *Se define la oscilación de la función f en el punto b como el límite* $\lim_{n \rightsquigarrow \infty} O_{I_n}$*, en donde (I_n) es una sucesión de intervalos encajados de la forma $[a_n,b]$, con $a \leq a_n < b$, y tales que $\bigcap_{n=1}^{\infty} I_n = \{b\}$.*

TEOREMA 37 (**Primer criterio de integrabilidad de Riemann**). *Sea $f : [a,b] \mapsto \mathbb{R}$ una función acotada, entonces f es Riemann integrable en el intervalo $[a,b]$ si y sólo si, para cualquier $\varepsilon > 0$, existe una partición P_ε del intervalo $[a,b]$ tal que si $P = \{x_0, \dots, x_n\}$ es un refinamiento de P_ε entonces:*

$$\sum_{i=1}^{n} O_{[x_{i-1},x_i]}(x_i - x_{i-1}) < \varepsilon$$

Demostración

Si f es Riemann integrable en el intervalo $[a,b]$, para cualquier $\varepsilon > 0$, existe una partición P_ε del intervalo $[a,b]$ tal que si P es refinamientos de P_ε entonces:

$|S(P;f) - S'(P;f)| < \frac{\varepsilon}{2}$

para cualquier par de sumas de Riemann, $S(P;f)$ y $S'(P;f)$, de f con respecto a P.

Como $O_{[x_{i-1},x_i]} = \sup\{f(x) - f(y) : x,y \in [x_{i-1},x_i]\}$, podemos tomar $\xi_i, \eta_i \in [x_{i-1},x_i]$ tales que:

$f(\xi_i) - f(\eta_i) > O_{[x_{i-1},x_i]} - \frac{\varepsilon}{2(b-a)}$

Entonces:

$\sum_{i=1}^{n} O_{[x_{i-1},x_i]}(x_i - x_{i-1}) < \sum_{i=1}^{n} \left[f(\xi_i) - f(\eta_i) + \frac{\varepsilon}{2(b-a)}\right](x_i - x_{i-1})$

$= \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) - \sum_{i=1}^{n} f(\eta_i)(x_i - x_{i-1}) + \frac{\varepsilon}{2} < \varepsilon$

Inversamente, supongamos que, para cualquier $\varepsilon > 0$, existe una partición P_ε del intervalo $[a,b]$ tal que si P es un refinamiento de P_ε entonces:

$\sum_{i=1}^{n} O_{[x_{i-1},x_i]}(x_i - x_{i-1}) < \frac{\varepsilon}{2}$

Sea $P_\varepsilon = \{y_0, \ldots, y_m\}$ y $P = \{x_0, \ldots, x_n\}$ un refinamiento de P_ε y consideremos dos sumas de Riemann, $S(P_\varepsilon; f)$ y $S(P; f)$ de f con respecto a P_ε y P, respectivamente:

$S(P_\varepsilon; f) = \sum_{i=1}^m f(\xi_i)(y_i - y_{i-1})$

$S(P; f) = \sum_{k=1}^n f(\eta_k)(x_k - x_{k-1})$

Como cada subintervalo de P_ε es unión de subintervalos de P, se tiene, para cada $i \in \{1, \ldots, m\}$:

$[y_{i-1}, y_i] = \bigcup_j \left[x_{k_j-1}, x_{k_j}\right]$

$\left| f(\xi_i)(y_i - y_{i-1}) - \sum_j f\left(\eta_{k_j}\right)\left(x_{k_j} - x_{k_j-1}\right)\right|$

$= \sum_j \left| f\left(\eta_{k_j}\right) - f(\xi_i)\right| \left(x_{k_j} - x_{k_j-1}\right) \leq O_{[y_{i-1}, y_i]} \sum_j \left(x_{k_j} - x_{k_j-1}\right)$

$= O_{[y_{i-1}, y_i]}(y_i - y_{i-1})$

Por lo tanto:

$|S(P_\varepsilon; f) - S(P; f)| \leq \sum_{i=1}^m O_{[y_{i-1}, y_i]}(y_i - y_{i-1}) < \frac{\varepsilon}{2}$

Así que, si P y P' son refinamientos de P_ε entonces:

$|S(P; f) - S(P'; f)| \leq |S(P_\varepsilon; f) - S(P; f)| + |S(P_\varepsilon; f) - S(P'; f)| < \varepsilon$

■

Teorema 38 (**Segundo criterio de integrabilidad de Riemann**). *Sea* $f : [a, b] \mapsto \mathbb{R}$ *una función acotada, entonces* f *es Riemann integrable en el intervalo* $[a, b]$ *si y sólo si, para cualquier* $\sigma > 0$ *y cualquier* $\varepsilon > 0$*, existe una partición* P_ε *del intervalo* $[a, b]$ *tal que si* $P = \{x_0, \ldots, x_n\}$ *es un refinamiento de* P_ε *entonces* $\lambda(P, \sigma) < \varepsilon$*, en donde* $\lambda(P, \sigma)$ *es la suma de las longitudes de los subintervalos de la partición* P *en los cuales la oscilación de la función es mayor o igual que* σ*.*

Demostración

Sea $\sigma > 0$ y $P = \{x_0, \ldots, x_n\}$ una partición del intervalo $[a, b]$, entonces:

$\sigma\lambda(P, \sigma) \leq \sum_{i=1}^n O_{[x_{i-1}, x_i]}(x_i - x_{i-1}) \leq O_{[a,b]}\lambda(P, \sigma) + (b - a)\sigma$

Supongamos que para cualquier $\sigma > 0$ y cualquier $\varepsilon > 0$, existe una partición P_ε del intervalo $[a, b]$ tal que si $P = \{x_0, \ldots, x_n\}$ es un refinamiento de P_ε entonces $\lambda(P, \sigma) < \frac{\varepsilon}{2O_{[a,b]}}$. Entonces, si P es un refinamiento de P_ε, se tiene:

$\sum_{i=1}^n O_{[x_{i-1}, x_i]}(x_i - x_{i-1}) \leq \frac{\varepsilon}{2} + (b - a)\sigma$

para cualquier $\sigma > 0$, por lo tanto:

$\sum_{i=1}^{n} O_{[x_{i-1},x_i]}(x_i - x_{i-1}) \leq \frac{\varepsilon}{2} < \varepsilon$

Así que se satisface el primer criterio de integrabilidad de Riemann.

Inversamente, supongamos que se satisface el primer criterio de integrabilidad de Riemann, entonces, para cualquier $\sigma > 0$ y cualquier $\varepsilon > 0$, existe una partición P_ε del intervalo $[a, b]$ tal que si $P = \{x_0, \ldots, x_n\}$ es un refinamiento de P_ε entonces:

$\sum_{i=1}^{n} O_{[x_{i-1},x_i]}(x_i - x_{i-1}) < \varepsilon\sigma$

Por lo tanto:

$\lambda(P, \sigma) \leq \frac{1}{\sigma} \sum_{i=1}^{n} O_{[x_{i-1},x_i]}(x_i - x_{i-1}) < \varepsilon$

■

TEOREMA 39 (**Caracterización de la funciones Riemann integrables**). *Sea* $f : [a, b] \mapsto \mathbb{R}$ *una función acotada, entonces* f *es Riemann integrable en el intervalo* $[a, b]$ *si y sólo si el conjunto de puntos en donde la función es discontinua tiene medida cero.*

Demostración

Para cada $\sigma > 0$, sea $D(\sigma)$ el conjunto de puntos $x \in [a, b]$ en los cuales la oscilación de f es mayor o igual que σ.

$D(\sigma)$ es un conjunto cerrado para cualquier $\sigma > 0$. En efecto, sea x un punto de acumulación de $D(\sigma)$, entonces toda vecindad de x contiene puntos de $D(\sigma)$, es decir, puntos en donde la oscilación de f es mayor o igual que σ. Por lo tanto, la oscilación de f en x es también mayor o igual que σ, así que x pertenece a $D(\sigma)$.

$D(\sigma)$ es, además, acotado. Por lo tanto $D(\sigma)$ es un conjunto compacto para cualquier $\sigma > 0$.

Por otra parte, si D es el conjunto de puntos en donde la función es discontinua, se tiene $D = \bigcup_{n=1}^{\infty} D(\frac{1}{n})$.

Dada $\sigma > 0$, si x es un punto en donde la oscilación de f es mayor o igual que σ y P es una partición del intervalo $[a, b]$, entonces, o bien x pertenece al interior de uno de los subintervalos de P en donde la oscilación de f es mayor o igual que σ, o bien x es un elemento de la partición P.

Si f es Riemann integrable, por el teorema 38, dada $\varepsilon > 0$, existe una partición P_ε del intervalo $[a, b]$ tal que si $P = \{x_0, \ldots, x_n\}$ es un refinamiento de P_ε entonces $\lambda(P, \sigma)$

$< \frac{\varepsilon}{2}$. Consideremos una familia de intervalos abiertos $I_1, I_2, \ldots, I_n$ cuya unión cubra los puntos de la partición P_ε y tales que la suma de sus longitudes sea menor que $\frac{\varepsilon}{2}$. De esta manera, $B(\sigma)$ queda cubierto por una colección finita de intervalos abiertos tales que la suma de sus longitudes es menor que ε, así que $B(\sigma)$ tiene medida cero. Por lo tanto, $D = \bigcup_{n=1}^{\infty} B(\frac{1}{n})$ también tiene medida cero.

Inversamente, supongamos que D tiene medida cero, entonces $B(\frac{1}{n})$ tiene medida cero para cualquier $n \in \mathbb{N}$. Dada $\sigma > 0$ arbitraria, tomemos $n > \frac{1}{\sigma}$, entonces $B(\sigma) \subset B(\frac{1}{n})$, así que $B(\sigma)$ tiene medida cero. Finalmente, $\{a, b\} \cup B(\sigma)$ también tiene medida cero.

Así que, dada $\varepsilon > 0$, existe una sucesión de intervalos abiertos no vacíos, $J_1, J_2, \ldots$, tales que:

(i) $\{a, b\} \cup B(\sigma) \subset \bigcup_k J_k$
(ii) $\sum_k l(J_k) < \varepsilon$

Como $\{a, b\} \cup B(\sigma)$ es compacto, existe una familia finita de intervalos J_k cuya unión también cubre a $\{a, b\} \cup B(\sigma)$. Sea $I_0 = (c_0, d_0), I_1 = (c_1, d_1), \ldots, I_{n+1} = (c_{n+1}, d_{n+1})$ esa familia de intervalos, en donde podemos asumir que:

$c_0 < a < d_0 \leq c_1 < d_1 \leq \cdots \leq c_n < d_n \leq c_{n+1} < b < d_{n+1}$

Se tiene entonces:

(i) $\{a, b\} \cup B(\sigma) \subset \cup_{k=0}^{n+1} I_k$
(ii) $\sum_{k=0}^{n+1} l\,(I_k) < \varepsilon$
(iii) En cada punto de un intervalo de la forma $[d_k, c_{k+1}]$ la oscilación de la función es menor que σ.

Dado un intervalo I de la forma $[d_k, c_{k+1}]$, si la oscilación de f en I es mayor o igual que σ, partamos I en dos subintervalos de la misma longitud; si, en ninguno de los subintervalos que se forman, la oscilación de f es mayor o igual que σ, terminemos el proceso; si no, partamos en dos subintervalos de la misma longitud cada subintervalo en donde la oscilación de f sea mayor o igual que σ así sucesivamente. Si este proceso no terminara nunca, obtendríamos un punto del intervalo $[d_k, c_{k+1}]$ en el cual la oscilación de la función es mayor o igual que σ, lo cual no es posible. Por lo tanto, continuando con ese proceso, en un número finito de pasos se tiene partido I de tal manera que en cada subintervalo de la partición la oscilación de f es menor que σ.

De esta manera se tiene una partición P_ε del intervalo $[a, b]$ formada por todos los extremos de los intervalos I_k y por todos los extremos de los subintervalos que conforman cada uno de los intervalos I de la forma $[d_k, c_{k+1}]$.

Sea P un refinamiento de P_ε. Si la oscilación de f es mayor o igual que σ en un subintervalo de P, entonces ese subintervalo está contenido en alguno de los intervalos I_k, por lo tanto, $\lambda(P,\sigma) < \varepsilon$. Así que, por el teorema 38, f es integrable.

■

COROLARIO 40. *Sea $f : [a,b] \mapsto \mathbb{R}$ una función continua, entonces f es Riemann integrable en el intervalo $[a,b]$.*

COROLARIO 41. *Sea $f : [a,b] \mapsto \mathbb{R}$ una función acotada y monótona no decreciente, entonces f es Riemann integrable en el intervalo $[a,b]$.*

COROLARIO 42. *Sea $f : [a,b] \mapsto \mathbb{R}$ una función acotada y monótona no creciente, entonces f es Riemann integrable en el intervalo $[a,b]$.*

TEOREMA 43. *Sean $f : [a,b] \mapsto \mathbb{R}$ y $g : [a,b] \mapsto \mathbb{R}$ dos funciones acotadas, Riemann integrables en el intervalo $[a,b]$, y tales que el conjunto $D = \{x \in [a,b] \mid f(x) \neq g(x)\}$ es finito o infinito numerable, entonces $\int_a^b f(x)dx = \int_a^b g(x)dx$.*

Demostración

Dada $\varepsilon > 0$ existen particiones P_ε y Q_ε del intervalo $[a,b]$ tales que si P es un refinamiento de P_ε y Q es un refinamiento de Q_ε, entonces $\left|S(P;f) - \int_a^b f(x)dx\right| < \frac{\varepsilon}{2}$ y $\left|S(Q;g) - \int_a^b g(x)dx\right| < \frac{\varepsilon}{2}$, para cualquier suma de Riemann, $S(P;f)$, de f con respecto a P y cualquier suma de Riemann, $S(Q;g)$, de g con respecto a Q.

Sea $P = \{x_0, \ldots, x_n\}$ un refinamiento de $P_\varepsilon \cup Q_\varepsilon$, entonces $\left|S(P;f) - \int_a^b f(x)dx\right| < \frac{\varepsilon}{2}$ y $\left|S(P;g) - \int_a^b g(x)dx\right| < \frac{\varepsilon}{2}$, para cualquier suma de Riemann, $S(P;f)$ de f con respecto a P y cualquier suma de Riemann, $S(P;g)$ de g con respecto a P.

Como el conjunto D es a lo más infinito numerable, podemos tomar dos sumas de Riemann $S(P;f) = \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1})$ y $S(P;g) = \sum_{i=1}^n g(\xi_i)(x_i - x_{i-1})$ de tal forma que los puntos ξ_i no pertenezcan a D, así que se tiene $S(P;f) = S(P;g)$. Por lo tanto:

$$\left|\int_a^b f(x)dx - \int_a^b g(x)dx\right| \leq \left|S(P;f) - \int_a^b f(x)dx\right| + \left|S(P;g) - \int_a^b g(x)dx\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Es decir, $\left|\int_a^b f(x)dx - \int_a^b g(x)dx\right| < \varepsilon$ para cualquier $\varepsilon > 0$, lo cual implica $\int_a^b f(x)dx = \int_a^b g(x)dx$.

■

LEMA 44. *Sea $A_1, A_2, \ldots$ una sucesión de subconjuntos Lebesgue medibles del intervalo (a,b) tales que $m(A_i) \geq \alpha$ para toda $i \in \mathbb{N}$, en donde α es un número real positivo fijo y m denota a la medida de Lebesgue. Entonces el conjunto:*

$$A = \{x \in (a,b) : x \in A_n \textit{ para una infinidad de valores de } n\}$$

es Lebesgue medible y $m(A) \geq \alpha$.

Demostración

Para cada $m \in \mathbb{N}$, sea $B_m = \bigcup_{n=m}^{\infty} A_n$. Entonces la sucesión de conjuntos B_m es monótona decreciente y $A = \bigcap_{m=1}^{\infty} B_m$, así que A es Lebesgue medible y:

$m(A) = m\left[\bigcap_{m=1}^{\infty} B_m\right] = \lim_{m\rightsquigarrow\infty} m\left[\bigcup_{n=m}^{\infty} A_n\right] \geq \alpha$

■

COROLARIO 45. *Sea* $A_1, A_2, \ldots$ *una sucesión de subconjuntos Lebesgue medibles del intervalo* (a,b) *tales que* $m(A_i) \geq \alpha$ *para toda* $i \in \mathbb{N}$, *en donde* α *es un número real positivo fijo. Entonces el conjunto:*

$$A = \{x \in (a,b) : x \in A_n \textit{ para una infinidad de valores de } n\}$$

es infinito no numerable.

Demostración

Si A fuera finito o infinito numerable se tendría $m(A) = 0$, lo cual contradice el lema 44.

■

TEOREMA 46. *Sea* $(f_n)_{n\in\mathbb{N}}$ *una sucesión monótona no decreciente de funciones no negativas, Riemann integrables en el intervalo* $[a,b]$, *y tales que la función* $f = \lim_{n\rightsquigarrow\infty} f_n$ *es también Riemann integrable en el intervalo* $[a,b]$, *entonces:*

$$\int_a^b f(x)dx = \lim_{n\rightsquigarrow\infty} \int_a^b f_n(x)dx$$

Demostración

Para cada $n \in \mathbb{N}$, sea $g_n = f - f_n$ y supongamos $\lim_{n\rightsquigarrow\infty} \frac{1}{b-a}\int_a^b g_n(x)dx = \alpha > 0$, entonces $\alpha_n = \int_a^b g_n(x)dx \geq \alpha(b-a) > 0$ para cualquier $n \in \mathbb{N}$.

Sea $P_0 = \{a, b\}$ y, para $n \in \mathbb{N}$, sea P_n una partición del intervalo $[a,b]$, que sea un refinamiento de P_{n-1} y tal que si $S(P_n; g_n) = \sum_{i=1}^{n} g_n(\xi_i)(x_i - x_{i-1})$ es cualquier suma de Riemann correspondiente a P_n, entonces $|S(P_n; g_n) - \alpha_n| < \frac{1}{3}\alpha(b-a)$.

Sea $E_n = \{x \in [a,b] \mid g_n(x) \geq \frac{1}{3}\alpha\}$, $\mathcal{A}_n$ la familia de subintervalos de la partición P_n que están contenidos en E_n, $\mathcal{B}_n$ la familia de subintervalos de la partición P_n que no están contenidos en E_n y L_n la suma de las longitudes de los intervalos de $\mathcal{A}_n$.

En cada subintervalo $[x_{i-1}, x_i]$ de $\mathcal{B}_n$ escojamos un punto ξ_i tal que $g_n(\xi_i) < \frac{1}{3}\alpha$, entonces:

$\frac{2}{3}\alpha(b-a) \leq \alpha_n - \frac{1}{3}\alpha(b-a) < S(P_n; g_n) = \sum_{i=1}^{n} g_n(\xi_i)(x_i - x_{i-1}) \leq \frac{1}{3}\alpha(b-a) + ML_n$

en donde $M = \sup\{f(x) \mid x \in [a,b]\}$.

Por lo tanto:

$L_n > \frac{1}{3M}\alpha(b-a)$

Así que, por el corolario 45, el conjunto:

$A = \{x \in [a,b] : x \in E_n$ para una infinidad de valores de $n\}$

es infinito no numerable, lo cual es imposible pues $\text{lím}_{n \rightsquigarrow \infty} g_n(x) = 0$ para cualquier $x \in [a,b]$.

Se concluye entonces que la hipótesis $\text{lím}_{n \rightsquigarrow \infty} \frac{1}{b-a} \int_a^b g_n(x)dx = \alpha > 0$ es falsa, es decir:

$\text{lím}_{n \rightsquigarrow \infty} \frac{1}{b-a} \int_a^b g_n(x)dx = 0$

de lo cual se sigue el resultado.

■

Teorema 47. *Sea $(f_n)_{n \in \mathbb{N}}$ una sucesión monótona no decreciente de funciones no negativas, Riemann integrables en todo intervalo finito, y tales que la función $f = \text{lím}_{n \rightsquigarrow \infty} f_n$ es también Riemann integrable en todo intervalo finito, entonces $\int_0^\infty f(x)dx = \text{lím}_{n \rightsquigarrow \infty} \int_0^\infty f_n(x)dx$.*

Demostración

Definamos $f_0 = 0$ y, para cada $j, k \in \mathbb{N}$, sea $a_{jk} = \int_{k-1}^{k} [f_j(x) - f_{j-1}(x)]\, dx$. Obsérvese que se tiene:

$\sum_{k=1}^\infty \sum_{j=1}^\infty a_{jk} = \int_0^\infty f(x)dx$ y $\sum_{j=1}^\infty \sum_{k=1}^\infty a_{jk} = \text{lím}_{n \rightsquigarrow \infty} \int_0^\infty f_n(x)dx$

En efecto, utilizando el teorema 46, se tiene:

$\sum_{k=1}^\infty \sum_{j=1}^\infty a_{jk} = \text{lím}_{m \rightsquigarrow \infty} \sum_{k=1}^m \sum_{j=1}^\infty a_{jk} = \text{lím}_{m \rightsquigarrow \infty} \sum_{k=1}^m \left(\text{lím}_{n \rightsquigarrow \infty} \sum_{j=1}^n a_{jk}\right)$

$= \text{lím}_{m \rightsquigarrow \infty} \sum_{k=1}^m \left(\text{lím}_{n \rightsquigarrow \infty} \int_{k-1}^k f_n(x)dx\right) = \text{lím}_{m \rightsquigarrow \infty} \sum_{k=1}^m \left(\int_{k-1}^k f(x)dx\right)$

$= \text{lím}_{m \rightsquigarrow \infty} \int_0^m f(x)dx = \int_0^\infty f(x)dx$

$\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk} = \text{lím}_{n\rightsquigarrow\infty}\sum_{j=1}^{n}\sum_{k=1}^{\infty} a_{jk} = \text{lím}_{n\rightsquigarrow\infty}\sum_{j=1}^{n}(\text{lím}_{m\rightsquigarrow\infty}\sum_{k=1}^{m} a_{jk})$

$= \text{lím}_{n\rightsquigarrow\infty}\sum_{j=1}^{n}\left(\text{lím}_{m\rightsquigarrow\infty}\int_0^m [f_j(x) - f_{j-1}(x)]\,dx\right)$

$= \text{lím}_{n\rightsquigarrow\infty}\left(\text{lím}_{m\rightsquigarrow\infty}\sum_{j=1}^{n}\int_0^m [f_j(x) - f_{j-1}(x)]\,dx\right)$

$= \text{lím}_{n\rightsquigarrow\infty}\left(\text{lím}_{m\rightsquigarrow\infty}\int_0^m f_n(x)dx\right) = \text{lím}_{n\rightsquigarrow\infty}\int_0^{\infty} f_n(x)dx$

El resultado se sigue entonces del lema 23.

■

A.4. La integral de Lebesgue

Recordemos que un conjunto $A \subset \mathbb{R}$ es Lebesgue medible si pertenece a la σ-álgebra de subconjuntos de $\mathbb{R}$ generada por los borelianos y los conjuntos de medida cero.

Recordemos también que la medida de Lebesgue m en $\mathbb{R}$ es la única función no negativa y σ-aditiva definida sobre los conjuntos Lebesgue medibles tal que $m(I)$ es igual a la longitud de I para cualquier intervalo I.

Consideraremos en esta parte funciones con valores en los reales extendidos, es decir, funciones que pueden tomar los valores ∞ y $-\infty$. Para $c \in \mathbb{R}$, haremos las convenciones siguientes.

$(-1)(\infty) = c - \infty = -\infty$

$(\infty)(\infty) = \infty + \infty = c + \infty = \infty$

$c(\infty) = \infty$ si $c > 0$

$(0)(\infty) = 0$

$\infty - \infty$ no está definido

$-\infty < c < \infty$

A.4.1. Funciones medibles.

Definición 48. *Sea $f : \mathbb{R} \mapsto \overline{\mathbb{R}}$. Diremos que f es medible si $\{x \in \mathbb{R} : f(x) \le y\}$ es Lebesgue medible, para cualquier $y \in \mathbb{R}$.*

Obsérvese que si $f : \mathbb{R} \mapsto \overline{\mathbb{R}}$ es medible, entonces los conjuntos $\{x \in \mathbb{R} : f(x) = -\infty\}$ y $\{x \in \mathbb{R} : f(x) = \infty\}$ son medibles. En efecto:

$\{x \in \mathbb{R} : f(x) = -\infty\} = \bigcap_{n=1}^{\infty} \{x \in \mathbb{R} : f(x) \leq -n\}$

$\{x \in \mathbb{R} : f(x) = \infty\}^c = \bigcup_{n=1}^{\infty} \{x \in \mathbb{R} : f(x) \leq n\}$

Los siguientes dos resultados se prueban fácilmente:

Proposición 49. *Una función $f : \mathbb{R} \mapsto \overline{\mathbb{R}}$ es medible si y sólo si se satisface cualquiera de las siguientes condiciones:*

$\{x \in \mathbb{R} : f(x) \leq y\}$ *es Lebesgue medible, para cualquier $y \in \mathbb{R}$.*

$\{x \in \mathbb{R} : f(x) < y\}$ *es Lebesgue medible, para cualquier $y \in \mathbb{R}$.*

$\{x \in \mathbb{R} : f(x) > y\}$ *es Lebesgue medible, para cualquier $y \in \mathbb{R}$.*

$\{x \in \mathbb{R} : f(x) \geq y\}$ *es Lebesgue medible, para cualquier $y \in \mathbb{R}$.*

Proposición 50. *Sea $f : \mathbb{R} \mapsto \overline{\mathbb{R}}$ una función medible, entonces $\{x \in \mathbb{R} : f(x) = \alpha\}$ y $f^{-1}(B)$ son Lebesgue medibles para cualquier $\alpha \in \overline{\mathbb{R}}$ y cualquier conjunto boreliano B.*

Obsérvese que toda función de la forma $\varphi = \sum_{k=1}^{m} b_k I_{F_k}$, en donde $b_1, \ldots, b_m$ son números reales y $F_1, \ldots, F_m$ son conjuntos Lebesgue medibles, es medible.

Definición 51. *Diremos que una función medible es simple si tiene la forma $\varphi = \sum_{k=1}^{m} b_k I_{F_k}$, en donde $b_1, \ldots, b_m$ son números reales y $F_1, \ldots, F_m$ son conjuntos Lebesgue medibles.*

Proposición 52. *Sea f una función medible no negativa, entonces existe una sucesión no decreciente de funciones simples no negativas φ_n tales que $\lim_{n \rightsquigarrow \infty} \varphi_n(x) = f(x)$ para cualquier $x \in \mathbb{R}$.*

Demostración

Para cada $n \in \mathbb{N}$, sea:

$$\varphi_n(x) = \begin{cases} \sum_{m=0}^{n2^n} \frac{m-1}{2^n} I_{\left\{y \in \mathbb{R} : \frac{m-1}{2^n} \leq f(y) < \frac{m}{2^n}\right\}}(x) & \text{si } f(x) < n \\ 0 & \text{si } f(x) \geq n \end{cases}$$

Dados $n \in \mathbb{N}$ y $x \in \mathbb{R}$, sea m el único número entero no negativo tal que $\frac{m-1}{2^n} \leq f(x) < \frac{m}{2^n}$. Entonces, como $\varphi_n(x) = \frac{m-1}{2^n}$, se tiene $f(x) - \frac{1}{2^n} < \varphi_n(x) \leq f(x)$, así que $\lim_{n \rightsquigarrow \infty} \varphi_n(x) = f(x)$.

Ahora bien, como $\frac{2(m-1)}{2^{n+1}} \leq f(x) < \frac{2m}{2^{n+1}}$, se tiene que $\frac{2m-2}{2^{n+1}} \leq f(x) < \frac{2m-1}{2^{n+1}}$ o bien $\frac{2m-1}{2^{n+1}} \leq f(x) < \frac{2m}{2^{n+1}}$. En el primer caso, se tiene $\varphi_{n+1}(x) = \frac{2m-2}{2^{n+1}} = \frac{m-1}{2^n} = \varphi_n(x)$, mientras que en el segundo, se tiene $\varphi_{n+1}(x) = \frac{2m-1}{2^{n+1}} > \frac{2m-2}{2^{n+1}} = \varphi_n(x)$. Así que, en cualquier caso, $\varphi_n(x) \leq \varphi_{n+1}(x)$.

Así que, φ_n es una sucesión monótona no decreciente de funciones simples no negativas tales que $\lim_{n\rightsquigarrow\infty} \varphi_n(x) = f(x)$ para cualquier $x \in \mathbb{R}$. ∎

PROPOSICIÓN 53. *Sea* $g_1 : \mathbb{R} \mapsto \overline{\mathbb{R}}$, $g_2 : \mathbb{R} \mapsto \overline{\mathbb{R}}$, ... *una sucesión de funciones medibles, entonces:*

(i) *Para cualquier* $n \in \mathbb{N}$, *las funciones* $\min\{g_1, \ldots, g_n\}$ *y* $\max\{g_1, \ldots, g_n\}$ *son medibles.*
(ii) *Las funciones* $\inf\{g_1, g_2, \ldots\}$ *y* $\sup\{g_1, g_2, \ldots\}$ *son medibles.*

Demostración

Sea $\mathcal{L}$ la σ-álgebra de los conjuntos Lebesgue medibles, entonces, para cualquier $y \in \mathbb{R}$, se tiene:

$\{x \in \mathbb{R} : \min\{g_1, \ldots, g_n\}(x) \geq y\} = \bigcap_{k=1}^{n} \{x \in \mathbb{R} : g_k(x) \geq y\} \in \mathcal{L}$

$\{x \in \mathbb{R} : \max\{g_1, \ldots, g_n\}(x) \leq y\} = \bigcap_{k=1}^{n} \{x \in \mathbb{R} : g_k(x) \leq y\} \in \mathcal{L}$

$\{x \in \mathbb{R} : \inf\{g_1, g_2, \ldots\}(x) \geq y\} = \bigcap_{k=1}^{\infty} \{x \in \mathbb{R} : g_k(x) \geq y\} \in \mathcal{L}$

$\{x \in \mathbb{R} : \sup\{g_1, g_2, \ldots\}(x) \leq y\} = \bigcap_{k=1}^{\infty} \{x \in \mathbb{R} : g_k(x) \leq y\} \in \mathcal{L}$

de lo cual se sigue el resultado.

∎

COROLARIO 54. *Sea* $g_1 : \mathbb{R} \mapsto \overline{\mathbb{R}}$, $g_2 : \mathbb{R} \mapsto \overline{\mathbb{R}}$, ... *una sucesión de funciones medibles, entonces las funciones* $\liminf g_n$ *y* $\limsup g_n$ *son medibles.*

Demostración

La sucesión $g_n = \inf\{f_j : j \geq n\}$ es monótona no decreciente y:

$\liminf_{n\rightsquigarrow\infty} f_n = \lim_{n\rightsquigarrow\infty} f_n = \inf f_n$

Así que $\liminf_{n\rightsquigarrow\infty} f_n$ es medible.

La sucesión $h_n = \sup\{f_j : j \geq n\}$ es monótona no creciente y:

$\limsup f_n = \lim_{n\rightsquigarrow\infty} f_n = \sup f_n$

Así que $\limsup f_n$ es medible.

∎

Corolario 55. *Sea* $g_1 : \mathbb{R} \mapsto \overline{\mathbb{R}}$, $g_2 : \mathbb{R} \mapsto \overline{\mathbb{R}}$, . . . *una sucesión de funciones medibles tales que* $\lim_{n \rightsquigarrow \infty} g_n(x)$ *existe para cualquier* $x \in \mathbb{R}$, *entonces la función* $g : \mathbb{R} \mapsto \overline{\mathbb{R}}$ *definida por* $g(x) = \lim_{n \rightsquigarrow \infty} g_n(x)$ *es medible.*

Lema 56. *Si una función* $f : \mathbb{R} \mapsto \overline{\mathbb{R}}$ *es medible, entonces* f^+ *y* f^- *son medibles.*

Demostración

La función idénticamente cero es medible, así que entonces $f^+ = \max\{f, 0\}$, es medible.

Para cualquier $y \in \mathbb{R}$, se tiene $\{x \in \mathbb{R} : [-f](x) \leq y\} = \{x \in \mathbb{R} : f(x) \geq y\}$, así que la función $-f$ es medible. Por lo tanto, $f^- = \max\{-f, 0\}$ es medible.

■

Proposición 57. *Sean* $f : \mathbb{R} \mapsto \mathbb{R}$ *y* $g : \mathbb{R} \mapsto \mathbb{R}$ *dos funciones medibles y* $c \in \mathbb{R}$. *Entonces las funciones* $f + c$, cf, $f + g$ *y* fg *son medibles.*

Demostración

Sean φ_n, Φ_n, ϕ_n y Ψ_n sucesiones no decrecientes de funciones simples no negativas tales que $\lim_{n \rightsquigarrow \infty} \varphi_n(x) = f^+(x)$, $\lim_{n \rightsquigarrow \infty} \Phi_n(x) = f^-(x)$, $\lim_{n \rightsquigarrow \infty} \phi_n(x) = g^+(x)$ y $\lim_{n \rightsquigarrow \infty} \Psi_n(x) = g^-(x)$, respectivamente, para cualquier $x \in \mathbb{R}$.

Las funciones $\varphi_n - \Phi_n + c$ y $c\varphi_n - c\Phi_n$ son simples y se tiene:

$\lim_{n \rightsquigarrow \infty} [\varphi_n - \Phi_n + c](x) = f(x) + c$

$\lim_{n \rightsquigarrow \infty} [c\varphi_n - c\Phi_n](x) = cf(x)$

Así que $f + c$ y cf son medibles.

También, las funciones $\varphi_n - \Phi_n + \phi_n - \Psi_n$ y $\varphi_n\phi_n + \Phi_n\Psi_n - \varphi_n\Psi_n - \Phi_n\phi_n$ son funciones simples y se tiene:

$\lim_{n \rightsquigarrow \infty} [\varphi_n - \Phi_n + \phi_n - \Psi_n](x) = [f + g](x)$

$\lim_{n \rightsquigarrow \infty} [\varphi_n\phi_n + \Phi_n\Psi_n - \varphi_n\Psi_n - \Phi_n\phi_n](x) = [fg](x)$

Así que $f + g$ y fg son medibles.

■

Proposición 58. *Sea* $f : \mathbb{R} \mapsto \overline{\mathbb{R}}$ *una función medible y* $g : \mathbb{R} \mapsto \overline{\mathbb{R}}$ *una función tal que* $g = f$ *excepto a lo más en un conjunto de medida cero, entonces* g *es medible.*

Demostración

Sea $E = \{x \in \mathbb{R} : g(x) = f(x)\}$, entonces, para cualquier $y \in \mathbb{R}$, se tiene.

$\{x \in \mathbb{R} : g(x) \leq y\}$

$[\{x \in \mathbb{R} : g(x) \leq y\} \cap E] \cup [\{x \in \mathbb{R} : g(x) \leq y\} \cap E^c]$

$= [\{x \in \mathbb{R} : f(x) \leq y\} \cap E] \cup \{x \in E^c : g(x) \leq y\}$

El conjunto $\{x \in E^c : g(x) \leq y\}$ tiene medida cero, por lo tanto es medible. Así que $\{x \in \mathbb{R} : g(x) \leq y\}$ es medible.

■

A.4.2. La integral de funciones medibles no negativas. Si $\varphi = \sum_{k=1}^{m} b_k I_{F_k}$ es una función simple entonces el conjunto de los valores que toma es finito. Sea $\{a_1, \ldots, a_n\}$ el conjunto formado por todos los distintos posibles valores no nulos de φ y, para $k \in \{1, \ldots, n\}$, sea $E_k = \{x \in R : \varphi(x) = a_k\}$, entonces los conjuntos $E_1, \ldots, E_n$ son ajenos por parejas y $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$. Esta última sumatoria será llamada la **representación canónica** de φ.

DEFINICIÓN 59. *Si φ es una función simple no negativa con representación canónica $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$, se define la integral de φ, $\int_{\mathbb{R}} \varphi(x)dx$, de la siguiente manera:*

$$\int_{\mathbb{R}} \varphi(x)dx = \sum_{k=1}^{n} a_k m(E_k)$$

LEMA 60. *Sea $\varphi = \sum_{j=1}^{m} b_j I_{F_j}$ una función simple no negativa, en donde los conjuntos $F_1, \ldots, F_m$ son ajenos por parejas, con representación canónica $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$, entonces $\sum_{j=1}^{m} b_j m(F_j) = \sum_{k=1}^{n} a_k m(E_k)$.*

Demostración

Para $k \in \{1, \ldots, n\}$, se tiene $E_k = \cup_{\{j \in \{1,\ldots,m\} : b_j = a_k\}} F_j$, así que:

$\sum_{k=1}^{n} a_k m(E_k) = \sum_{k=1}^{n} a_k \sum_{\{j \in \{1,\ldots,m\} : b_j = a_k\}} m(F_j)$

$= \sum_{k=1}^{n} \sum_{\{j \in \{1,\ldots,m\} : b_j = a_k\}} b_j m(F_j) = \sum_{j=1}^{m} b_j m(F_j)$

■

PROPOSICIÓN 61. *Sea $\varphi = \sum_{j=1}^{m} b_j I_{F_j}$ una función simple, en donde los coeficientes $b_1, \ldots, b_m$ son no negativos, con representación canónica $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$, entonces $\sum_{j=1}^{m} b_j m(F_j) = \sum_{k=1}^{n} a_k m(E_k)$.*

Demostración

Los términos de la sumatoria $\sum_{j=1}^{m} b_j I_{F_j}$ en los cuales $b_j = 0$ pueden eliminarse, así que podemos asumir que $b_1, \ldots, b_m$ son números reales positivos.

Sea $F = \cup_{j=1}^{m} F_m$, $T = \{1, \ldots, m\}$ y, si $A = \{i_1, \ldots, i_k\} \subset T$ y $T - A = \{j_1, \ldots, j_{m-k}\}$, definamos:

$F_A = F \cap F_{i_1} \cap \cdots \cap F_{i_k} \cap F_{j_1}^c \cap \cdots \cap F_{j_{m-k}}^c$

$c_A = b_{i_1} + \cdots + b_{i_k}$

Entonces $F = \cup_{A \subset T} F_A$, $F_i = \cup_{\{A \subset T : i \in A\}} F_A$ y, si A y B son dos subconjuntos distintos de T, F_A y F_B son ajenos. Por lo tanto:

$m(F_j) = \sum_{\{A \subset T : j \in A\}} m(F_A)$

Así que:

$\sum_{j=1}^{m} b_j m(F_j) = \sum_{j=1}^{m} b_j \sum_{\{A \subset T : j \in A\}} m(F_A)$

$= \sum_{j=1}^{m} \sum_{\{A \subset T : j \in A\}} b_j m(F_A) = \sum_{A \subset T} \sum_{j \in A} b_j m(F_A)$

$= \sum_{A \subset T} c_A m(F_A)$

Además, si $x \in F_A$ y $A = \{i_1, \ldots, i_k\}$, $\varphi(x) = b_{i_1} + \cdots + b_{i_k} = c_A$, por lo tanto $\varphi = \sum_{A \subset T} c_A I_{F_A}$.

Así que se tiene:

$\sum_{j=1}^{m} b_j I_{F_j} = \sum_{A \subset T} c_A I_{F_A}$

$\sum_{j=1}^{m} b_j m(F_j) = \sum_{A \subset T} c_A m(F_A)$

Pero como $\sum_{i=1}^{n} a_i I_{E_i} = \sum_{j=1}^{m} b_j I_{F_j}$, se tiene $\sum_{i=1}^{n} a_i I_{E_i} = \sum_{A \subset T} c_A I_{E_A}$, así que, por el lema anterior:

$\sum_{j=1}^{m} b_j m(F_j) = \sum_{A \subset T} c_A m(F_A) = \sum_{k=1}^{n} a_k m(E_k)$

■

COROLARIO 62. *Sea* $\varphi = \sum_{j=1}^{m} b_j I_{F_j}$ *una función simple, en donde los coeficientes* $b_1, \ldots, b_m$ *son no negativos, entonces:*

$$\int_{\mathbb{R}} \varphi(x) dx = \sum_{j=1}^{m} b_j m(F_j)$$

PROPOSICIÓN 63. *Sean* φ *y* ψ *dos funciones simples no negativas, entonces:*

(i) $\int_{\mathbb{R}}[a\varphi(x)+b\psi(x)]\,dx = a\int_{\mathbb{R}}\varphi(x)dx + b\int_{\mathbb{R}}\psi(x)dx$ *para cualesquiera números reales a y b no negativos.*
(ii) *Si $\varphi \leq \psi$, entonces $\int_{\mathbb{R}}\varphi(x)dx \leq \int_{\mathbb{R}}\psi(x)dx$.*

Demostración

i) Sean $\sum_{k=1}^{n} a_k I_{E_k}$ y $\sum_{k=1}^{m} b_k I_{F_k}$ las representaciones canónicas de φ y ψ, respectivamente. Entonces:

$a\varphi + b\psi = \sum_{k=1}^{n} aa_k I_{E_k} + \sum_{k=1}^{m} bb_k I_{F_k}$

Así que:

$\int_{\mathbb{R}}[a\varphi(x)+b\psi(x)]\,dx = \sum_{k=1}^{n} aa_k m(E_k) + \sum_{k=1}^{m} bb_k m(F_k)$

$= a\int_{\mathbb{R}}\varphi(x)dx + b\int_{\mathbb{R}}\psi(x)dx$

ii) Si $\varphi \leq \psi$, entonces $\psi - \varphi$ es una función simple no negativa y $\psi = \varphi + (\psi - \varphi)$, así que:

$\int_{\mathbb{R}}\psi(x)dx = \int_{\mathbb{R}}\varphi(x)dx + \int_{\mathbb{R}}[\psi(x)-\varphi(x)]\,dx$

Por lo tanto:

$\int_{\mathbb{R}}\psi(x)dx - \int_{\mathbb{R}}\varphi(x)dx = \int_{\mathbb{R}}[\psi(x)-\varphi(x)]\,dx \geq 0$

■

Definición 64. *Si f es una función medible no negativa, se define la integral de f, $\int_{\mathbb{R}} f(x)dx$, de la siguiente manera:*

$$\int_{\mathbb{R}} f(x)dx = \sup\left\{\int_{\mathbb{R}}\varphi(x)dx : \varphi \text{ es simple y } 0 \leq \varphi \leq f\right\}$$

Definición 65. *Si f es una función medible no negativa y E es un conjunto Lebesgue medible, se define:*

$$\int_E f(x)dx = \int_{\mathbb{R}} I_E(x) f(x)dx$$

Lema 66. *Sea φ una función simple no negativa. Entonces, la función $m : \mathcal{L} \mapsto \mathbb{R}$, definida por $m(E) = \int_E \varphi(x)dx$, es una medida.*

Demostración

Obviamente, m es no negativa y, por la proposición anterior, es finitamente aditiva.

Sea $\sum_{k=1}^{n} a_k I_{E_k}$ la representación canónica de φ, A_n una sucesión monótona no decreciente de conjuntos lebesgue medibles y $A = \cup_{n=1}^{\infty} A_n$. Entonces

$\int_{A_n} \varphi(x)dx = \sum_{k=1}^{n} a_k m\left(A_n \cap E_k\right)$

$\int_{A} \varphi(x)dx = \sum_{k=1}^{n} a_k m\left(A \cap E_k\right)$

Así que:

$\text{lím}_{n \rightsquigarrow \infty} m(A_n) = \text{lím}_{n \rightsquigarrow \infty} \int_{A_n} \varphi(x)dx = \text{lím}_{n \rightsquigarrow \infty} \sum_{k=1}^{n} a_k m\left(A_n \cap E_k\right)$

$= \sum_{k=1}^{n} a_k \, \text{lím}_{n \rightsquigarrow \infty} m\left(A_n \cap E_k\right) = \sum_{k=1}^{n} a_k m\left(A \cap E_k\right)$

$= \int_{A} \varphi(x)dx = m(A)$

■

TEOREMA 67 (**Teorema de la convergencia monótona**). *Sea f_n una sucesión no decreciente de funciones medibles no negativas, entonces:*

$$\int_{\mathbb{R}} \text{lím}_{n \rightsquigarrow \infty} f_n(x)dx = \text{lím}_{n \rightsquigarrow \infty} \int_{\mathbb{R}} f_n(x)dx$$

Demostración

Sea $f = \text{lím}_{n \rightsquigarrow \infty} f_n$ y φ una función simple no negativa tal que $\varphi \leq f$, $\alpha \in (0,1)$ y $A_n = \{x \in \mathbb{R}: f_n(x) \geq \alpha\varphi(x)\}$. Entonces, la sucesión A_n es monótona no decreciente y $\cup A_n = \mathbb{R}$. Además, la función $m : \mathcal{L} \mapsto \mathbb{R}$, definida por $m(E) = \int_E \varphi(x)dx$, es una medida, así que:

$\text{lím}_{n \rightsquigarrow \infty} \int_{A_n} \varphi(x)dx = \int_{\mathbb{R}} \varphi(x)dx$

Por otra parte, $\alpha \int_{A_n} \varphi(x)dx \leq \int_{A_n} f_n(x)dx \leq \int_{\mathbb{R}} f_n(x)dx$ para cualquier $n \in \mathbb{N}$, así que:

$\alpha \int_{\mathbb{R}} \varphi(x)dx = \alpha \, \text{lím}_{n \rightsquigarrow \infty} \int_{A_n} \varphi(x)dx \leq \text{lím}_{n \rightsquigarrow \infty} \int_{\mathbb{R}} f_n(x)dx$

Haciendo tender α a 1, se obtiene entonces:

$\int_{\mathbb{R}} \varphi(x)dx \leq \text{lím}_{n \rightsquigarrow \infty} \int_{\mathbb{R}} f_n(x)dx$

Por lo tanto:

$\int_{\mathbb{R}} f(x)dx \leq \text{lím}_{n \rightsquigarrow \infty} \int_{\mathbb{R}} f_n(x)dx$

■

PROPOSICIÓN 68. *Sean f y g dos funciones medibles no negativas, entonces:*

(i) $\int_{\mathbb{R}} \left[af(x) + bg(x)\right] dx = a \int_{\mathbb{R}} f(x)dx + b \int_{\mathbb{R}} g(x)dx$ *para cualesquiera números reales a y b no negativos.*
(ii) *Si $f \leq g$, entonces $\int_{\mathbb{R}} f(x)dx \leq \int_{\mathbb{R}} g(x)dx$.*

Demostración

La segunda propiedad es inmediata de la definición. Para la primera, sean φ_n y ψ_n dos sucesiones no decrecientes de funciones simples no negativas tales que $\lim_{n\rightsquigarrow\infty}\varphi_n(x) = f(x)$ y $\lim_{n\rightsquigarrow\infty}\psi_n(x) = g(x)$ para cualquier $x \in \mathbb{R}$.

Para cada $n \in \mathbb{N}$, se tiene:

$\int_{\mathbb{R}} [a\varphi_n(x) + b\psi_n(x)]\, dx = a\int_{\mathbb{R}} \varphi_n(x)dx + b\int_{\mathbb{R}} \psi_n(x)dx$

Así que, por el teorema de la convergencia monótona, se tiene:

$\int_{\mathbb{R}} [af(x) + bg(x)]\, dx = a\int_{\mathbb{R}} f(x)dx + b\int_{\mathbb{R}} f_n(x)dx$

■

PROPOSICIÓN 69 (**Lema de Fatou**). *Sea $\{f_n\}_{n\in N}$ una sucesión de funciones medibles no negativas. Entonces:*

$$\int_{\mathbb{R}} \liminf_{n\rightsquigarrow\infty} f_n(x)dx \leq \liminf_{n\rightsquigarrow\infty} \int_{\mathbb{R}} f_n(x)dx$$

Demostración

La sucesión $g_n = \inf\{f_j : j \geq n\}$ es monótona no decreciente y $\liminf_{n\rightsquigarrow\infty} f_n = \lim_{n\rightsquigarrow\infty} f_n$, así que, por el teorema de la convergencia monótona:

$\int_{\mathbb{R}} \liminf_{n\rightsquigarrow\infty} f_n(x)dx = \lim_{n\rightsquigarrow\infty} \int_{\mathbb{R}} g_n(x)dx$

Por otra parte, $g_n \leq f_j$ para cualquier $j \geq n$, así que:

$\int_{\mathbb{R}} g_n(x)dx \leq \inf\left\{\int_{\mathbb{R}} f_j(x)dx : j \geq n\right\}$

Por lo tanto:

$\int_{\mathbb{R}} \liminf_{n\rightsquigarrow\infty} f_n(x)dx = \lim_{n\rightsquigarrow\infty} \int_{\mathbb{R}} g_n(x)dx$

$\leq \lim_{n\rightsquigarrow\infty} \inf\left\{\int_{\mathbb{R}} f_j(x)dx : j \geq n\right\} = \liminf_{n\rightsquigarrow\infty} \int_{\mathbb{R}} f_n(x)dx$

■

A.4.3. Funciones integrables.

DEFINICIÓN 70. *Se dice que una función medible f es integrable sobre un conjunto Lebesgue medible E si $\int_E |f(x)|\, dx < \infty$.*

PROPOSICIÓN 71. *Una función medible f es integrable sobre un conjunto Lebesgue medible E si y sólo si f^+ y f^- son integrables sobre E.*

Demostración

Se tiene $f^+ \leq |f|$ y $f^- \leq |f|$, así que si f es una función medible integrable sobre E, entonces f^+ y f^- son también integrables sobre E.

Inversamente, si f^+ y f^- son integrables sobre E, entonces $\int_E |f(x)|\, dx = \int_E f^+(x)dx + \int_E f^-(x)dx$, así que f es también integrable sobre E.

■

DEFINICIÓN 72. *Si f una función medible e integrable sobre un conjunto Lebesgue medible E, se define su integral sobre E, $\int_E f(x)dx$, de la siguiente manera:*

$$\int_E f(x)dx = \int_E f^+(x)dx - \int_E f^-(x)dx$$

PROPOSICIÓN 73. *Sean f y g dos funciones medibles e integrables sobre un conjunto Lebesgue medible E, entonces:*

(i) *Para cualesquiera números reales a y b, la función $af+bg$ es integrable sobre E y:*
$\int_{\mathbb{E}} [af(x) + bg(x)]\, dx = a \int_{\mathbb{E}} f(x)dx + b \int_{\mathbb{E}} g(x)dx$
(ii) *Si $f \leq g$ sobre E, entonces $\int_{\mathbb{E}} f(x)dx \leq \int_{\mathbb{E}} g(x)dx$.*
(iii) $\left|\int_E f(x)dx\right| \leq \int_E |f(x)|\, dx$

Demostración

i) $|af(x) + bg(x)| \leq |a|\,|f(x)| + |b|\,|g(x)|$, así que $\int_{\mathbb{E}} |af(x) + bg(x)|\, dx < \infty$

Si $a < 0$, se tiene:

$(af)^+ = |a|\, f^-$ y $(af)^- = |a|\, f^+$, así que:

$\int_E af(x)dx = \int_E |a|\, f^-(x)dx - \int_E |a|\, f^+(x)dx$

$= -a \int_E f^-(x)dx + a \int_E f^+(x)dx = a \left(\int_E f^-(x)dx + \int_E f^+(x)dx\right)$

$= a \int_E f(x)dx$

Si $a \geq 0$, se tiene:

$(af)^+ = af^+$y $(af)^- = af^-$, así que:

$\int_{\mathbb{E}} af(x)dx = \int_{\mathbb{E}} af^+(x)dx - \int_{\mathbb{E}} af^-(x)dx = a \left(\int_{\mathbb{E}} f^+(x)dx - \int_{\mathbb{E}} f^-(x)dx\right)$

$= a \int_{\mathbb{E}} f(x)dx$

Además:

$(f+g)^{+}-(f+g)^{-}=f+g=(f^{+}-f^{-})+(g^{+}-g^{-})=(f^{+}+g^{+})-(f^{-}+g^{-})$

Así que:

$(f+g)^{+}+f^{-}+g^{-}=(f+g)^{-}+f^{+}+g^{+}$

Por lo tanto:

$\int_{\mathbb{E}}(f+g)^{+}(x)dx+\int_{\mathbb{E}}f^{-}(x)dx+\int_{\mathbb{E}}g^{-}(x)dx = \int_{\mathbb{E}}(f+g)^{-}(x)dx+\int_{\mathbb{E}}f^{+}(x)dx+\int_{\mathbb{E}}g^{+}(x)dx$

De lo cual se sigue:

$\int_{\mathbb{E}}(f+g)(x)dx=\int_{\mathbb{E}}(f+g)^{+}(x)dx-\int_{\mathbb{E}}(f+g)^{-}(x)dx$

$=\int_{\mathbb{E}}f^{+}(x)dx+\int_{\mathbb{E}}g^{+}(x)dx-\int_{\mathbb{E}}f^{-}(x)dx-\int_{\mathbb{E}}g^{-}(x)dx$

$=\left(\int_{\mathbb{E}}f^{+}(x)dx-\int_{\mathbb{E}}f^{-}(x)dx\right)+\left(\int_{\mathbb{E}}g^{+}(x)dx-\int_{\mathbb{E}}g^{-}(x)dx\right)$

$=\int_{\mathbb{E}}f(x)dx+\int_{\mathbb{E}}g(x)dx$

ii) Si $f\leq g$ sobre E, entonces $g-f\geq 0$ sobre E, así que:

$\int_{\mathbb{E}}g(x)dx-\int_{\mathbb{E}}f(x)dx=\int_{\mathbb{E}}(g-f)(x)dx\geq 0$

Por lo tanto:

$\int_{\mathbb{E}}f(x)dx\leq\int_{\mathbb{E}}g(x)dx$

iii) $\left|\int_E f(x)dx\right|=\left|\int_E f^{+}(x)dx-\int_E f^{-}(x)dx\right|$

$\leq\int_E f^{+}(x)dx+\int_E f^{-}(x)dx=\int_E|f(x)|\,dx$

■

Proposición 74. *Sea $f:\mathbb{R}\mapsto\overline{\mathbb{R}}$ una función boreliana e integrable sobre un conjunto boreliano E y tal que $\int_A f(x)dx=0$ para cualquier conjunto boreliano $A\subset E$, entonces el conjunto $\{x\in E: f(x)\neq 0\}$ tiene medida cero.*

Demostración

Sea $B=\{x\in E: f(x)>0\}$ y, para cada $n\in\mathbb{N}$, sea $B_n=\left\{x\in E: f(x)>\frac{1}{n}\right\}$, entonces:

$0=\int_{B_n}f(x)dx\geq\frac{1}{n}m(B_n)$

Por lo tanto, $m(B_n)=0$ para cualquier $n\in\mathbb{N}$. Así que:

$m\left(B\right) = m\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\rightsquigarrow\infty} m\left(B_n\right) = 0$

De la misma manera, se tiene que $m\left(\{x \in E : f(x) < 0\}\right) = 0$. Así que:

$m\left(\{x \in E : f(x) \neq 0\}\right) = 0$

■

Corolario 75. *Sean f y g dos funciones borelianas e integrables sobre un conjunto boreliano E, y tales que $\int_A f(x)dx = \int_A g(x)dx$ para cualquier conjunto boreliano $A \subset E$, entonces el conjunto $\{x \in E : f(x) \neq g(x)\}$ tiene medida cero.*

Proposición 76. *Sea g una función no negativa, integrable sobre un conjunto Lebesgue medible E y $\{f_n\}_{n\in N}$ una sucesión de funciones medibles tales que $|f_n| \leq g$ y $f = \lim_{n\rightsquigarrow\infty} f_n$ existe c.s., entonces $\lim_{n\rightsquigarrow\infty} \int_E |f_n(x) - f(x)|\, dx = 0$.*

Demostración

Para cada $n \in \mathbb{N}$, sea $h_n = 2g - |f_n - f|$, entonces, por el lema de Fatou, se tiene:

$2\int_E f(x)dx = \int_E \lim_{n\rightsquigarrow\infty} h_n(x)dx$

$\leq \liminf_{n\rightsquigarrow\infty} \int_E h_n(x)dx = 2\int_E g(x)dx - \limsup_{n\rightsquigarrow\infty} \int_E |f_n(x) - f(x)|\, dx$

Así que:

$\limsup_{n\rightsquigarrow\infty} \int_E |f_n(x) - f(x)|\, dx = 0$

■

Corolario 77 (**Teorema de la convergencia dominada**). *Sea g una función no negativa integrable sobre un conjunto Lebesgue medible E y $\{f_n\}_{n\in N}$ una sucesión de funciones medibles tales que $|f_n| \leq g$ y $\lim_{n\rightsquigarrow\infty} f_n$ existe c.s., entonces:*

$$\int_E \lim_{n\rightsquigarrow\infty} f_n(x)dx = \lim_{n\rightsquigarrow\infty} \int_E f_n(x)dx$$

Proposición 78. *Sea $f : [a,b] \mapsto \mathbb{R}$ una función Riemann integrable no negativa, entonces f es medible y la integral de Lebesgue de f coincide con su integral de Riemann.*

Demostración

Sea $P_n = \{x_0^{(n)}, \ldots, x_{m_n}^{(n)}\}$ una sucesión de particiones del intervalo $[a,b]$ tales que, para cualquier $n \in \mathbb{N}$, P_{n+1} es un refinamiento de P_n y $\sum_{i=1}^{m_n} (M_i^{(n)} - m_i^{(n)})(x_i^{(n)} - x_{i-1}^{(n)}) < \frac{1}{n}$, en donde $M_i^{(n)} = \sup\{f(x) \mid x \in [x_{i-1}^{(n)}, x_i^{(n)}]\}$ y $m_i^{(n)} = \inf\{f(x) \mid x \in [x_{i-1}^{(n)}, x_i^{(n)}]\}$.

Sean:

$\varphi_n = \sum_{i=1}^{m_n-1} m_i^{(n)} I_{\left[x_{i-1}^{(n)}, x_i^{(n)}\right)} + m_{m_n}^{(n)} I_{\left[x_{m_n-1}^{(n)}, x_{m_n}^{(n)}\right]}$

$\psi_n = \sum_{i=1}^{m_n-1} M_i^{(n)} I_{\left[x_{i-1}^{(n)}, x_i^{(n)}\right)} + M_{m_n}^{(n)} I_{\left[x_{m_n-1}^{(n)}, x_{m_n}^{(n)}\right]}$

Entonces, denotando por $(R)\int$ a la integral de Riemann, se tiene que: la sucesión $(R)\int_a^b \varphi_n(x)dx$ es monótona creciente, la sucesión $(R)\int_a^b \psi_n(x)dx$ es monótona decreciente y:

$\lim_{n\rightsquigarrow\infty} (R)\int_a^b \varphi_n(x)dx = \lim_{n\rightsquigarrow\infty} (R)\int_a^b \psi_n(x)dx = (R)\int_a^b f(x)dx$

Por otra parte, para cualquier $n \in \mathbb{N}$, φ_n y ψ_n son funciones simples tales que $\varphi_n \leq f \leq \psi_n$, así que $\varphi^* = \sup \varphi_n$ y $\psi^* = \inf \psi_n$ son funciones medibles y $\varphi^* \leq f \leq \psi^*$. Además:

$\{x \in [a,b] : \psi^*(x) - \varphi^*(x) > 0\}$

$= \cup_{k=1}^{\infty} \left\{x \in [a,b] : \psi^*(x) - \varphi^*(x) > \frac{1}{k}\right\}$

Pero, para cualquier $n \in \mathbb{N}$, se tiene:

$\frac{1}{k} m\left(\left\{x \in [a,b] : \psi^*(x) - \varphi^*(x) > \frac{1}{k}\right\}\right)$

$\leq \frac{1}{k} m\left(\left\{x \in [a,b] : \psi_n(x) - \varphi_n(x) > \frac{1}{k}\right\}\right)$

$\leq \int_{\left\{x \in [a,b] : \psi_n(x) - \varphi_n(x) > \frac{1}{k}\right\}} \left[\psi_n(x) - \varphi_n(x)\right] dx$

$\leq \int_a^b \left[\psi_n(x) - \varphi_n(x)\right] dx < \frac{1}{n}$

Así que:

$m\left(\left\{x \in [a,b] : \psi^*(x) - \varphi^*(x) > \frac{1}{k}\right\}\right) < \frac{k}{n}$

para cualquier $n \in \mathbb{N}$. Por lo tanto:

$m\left(\left\{x \in [a,b] : \psi^*(x) - \varphi^*(x) > \frac{1}{k}\right\}\right) = 0$

Por consiguiente:

$m\left(\{x \in [a,b] : \psi^*(x) - \varphi^*(x) > 0\}\right)$

$\leq \sum_{k=1}^{\infty} m\left(\left\{x \in [a,b] : \psi^*(x) - \varphi^*(x) > \frac{1}{k}\right\}\right) = 0$

Se sigue entonces que $\varphi^* = f = \psi^*$ excepto a lo más en un conjunto de medida cero, así que f es medible.

Por otra parte, para cualquier $n \in \mathbb{N}$, se tiene:

$\varphi_n \leq \varphi^* \leq f \leq \psi^* \leq \psi_n$

Así que:

$\int_a^b \varphi_n(x)dx \leq \int_a^b \varphi^*(x)dx \leq \int_a^b f(x)dx \leq \int_a^b \psi^*(x)dx \leq \int_a^b \psi_n(x)dx$

Pero :

$\int_a^b \varphi_n(x)dx = (R)\int_a^b \varphi_n(x)dx$

$\int_a^b \psi_n(x)dx = (R)\int_a^b \psi_n(x)dx$

Por lo tanto:

$(R)\int_a^b \varphi_n(x)dx \leq \int_a^b \varphi^*(x)dx \leq \int_a^b f(x)dx \leq \int_a^b \psi^*(x)dx \leq (R)\int_a^b \psi_n(x)dx$

Así que, tomando límites cuando $n \rightsquigarrow \infty$, se tiene:

$(R)\int_a^b f(x)dx \leq \int_a^b \varphi^*(x)dx \leq \int_a^b f(x)dx \leq \int_a^b \psi^*(x)dx \leq (R)\int_a^b f(x)dx$

De lo cual se sigue:

$\int_a^b f(x)dx = (R)\int_a^b f(x)dx$

■

A.4.4. La integral de Lebesgue en $\mathbb{R}^2$. Las definiciones y resultados sobre la integral se extienden a $\mathbb{R}^n$, siendo las definiciones y demostraciones esencialmente las mismas. A continuación se enuncian para el caso de $\mathbb{R}^2$.

Recordemos que un conjunto $A \subset \mathbb{R}$ es Lebesgue medible si pertenece a la σ-álgebra de subconjuntos de $\mathbb{R}^2$ generada por los borelianos y los conjuntos de medida cero.

Recordemos también que la medida de Lebesgue m en $\mathbb{R}^2$ es la única función no negativa y σ-aditiva definida sobre los conjuntos Lebesgue medibles tal que $m(I \times J)$ es igual al producto de las longitudes de I y J, para cualquier par de intervalos I y J.

DEFINICIÓN 79. *Sea $f : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$. Diremos que f es medible si:*

$$\{(x, y) \in \mathbb{R}^2 : f\,(x, y) \leq y\}$$

es Lebesgue medible, para cualquier $x \in \mathbb{R}$.

Proposición 80. *Sea* $f : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$ *una función medible, entonces y* $f^{-1}(B)$ *es Lebesgue medible para cualquier conjunto boreliano* B.

Definición 81. *Diremos que una función medible es simple si tiene la forma* $\varphi = \sum_{k=1}^{m} b_k I_{F_k}$, *en donde* $b_1, \ldots, b_m$ *son números reales y* $F_1, \ldots, F_m$ *son conjuntos Lebesgue medibles de* $\mathbb{R}^2$.

Proposición 82. *Sea* f *una función medible no negativa, entonces existe una sucesión no decreciente de funciones simples no negativas* φ_n *tales que* $\lim_{n \rightsquigarrow \infty} \varphi_n(x, y) = f(x, y)$ *para cualquier* $(x, y) \in \mathbb{R}^2$.

Proposición 83. *Sea* $g_1 : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$, $g_2 : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$, $\ldots$ *una sucesión de funciones medibles, entonces:*

(i) *Para cualquier* $n \in \mathbb{N}$, *las funciones* $\min\{g_1, \ldots, g_n\}$ *y* $\max\{g_1, \ldots, g_n\}$ *son medibles.*
(ii) *Las funciones* $\inf\{g_1, g_2, \ldots\}$ *y* $\sup\{g_1, g_2, \ldots\}$ *son medibles.*

Corolario 84. *Sea* $g_1 : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$, $g_2 : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$, $\ldots$ *una sucesión de funciones medibles, entonces las funciones* $\liminf g_n$ *y* $\limsup g_n$ *son medibles.*

Corolario 85. *Sea* $g_1 : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$, $g_2 : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$, $\ldots$ *una sucesión de funciones medibles tales que* $\lim_{n \rightsquigarrow \infty} g_n(x)$ *existe para cualquier* $x \in \mathbb{R}$, *entonces la función* $g : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$ *definida por* $g(x, y) = \lim_{n \rightsquigarrow \infty} g_n(x, y)$ *es medible.*

Lema 86. *Si una función* $f : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$ *es medible, entonces* f^+ *y* f^- *son medibles.*

Proposición 87. *Sean* $f : \mathbb{R}^2 \mapsto \mathbb{R}$ *y* $g : \mathbb{R}^2 \mapsto \mathbb{R}$ *dos funciones medibles y* $c \in \mathbb{R}$. *Entonces las funciones* $f + c$, cf, $f + g$ *y* fg *son medibles.*

Proposición 88. *Sea* $f : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$ *una función medible y* $g : \mathbb{R}^2 \mapsto \mathbb{R}$ *una función tal que* $g = f$ *excepto a lo más en un conjunto de medida cero, entonces* g *es medible.*

Si $\varphi = \sum_{k=1}^{m} b_k I_{F_k}$ es una función simple entonces el conjunto de los valores que toma es finito. Sea $\{a_1, \ldots, a_n\}$ el conjunto formado por todos los distintos posibles valores no nulos de φ y, para $k \in \{1, \ldots, n\}$, sea $E_k = \{(x, y) \in \mathbb{R}^2 : \varphi(x, y) = a_k\}$, entonces los conjuntos $E_1, \ldots, E_n$ son ajenos por parejas y $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$. Esta última sumatoria será llamada la **representación canónica** de φ.

Definición 89. *Si* φ *es una función simple no negativa con representación canónica* $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$, *se define la integral de* φ, $\int_{\mathbb{R}^2} \varphi dm$, *de la siguiente manera:*

$$\int_{\mathbb{R}^2} \varphi dm = \sum_{k=1}^{n} a_k m(E_k)$$

Definición 90. *Si* f *es una función medible no negativa, se define la integral de* f, $\int_{\mathbb{R}^2} f dm$, *de la siguiente manera:*

$$\int_{\mathbb{R}^2} f dm = \sup \left\{ \int_{\mathbb{R}^2} \varphi dm : \varphi \text{ es simple y } 0 \leq \varphi \leq f \right\}$$

DEFINICIÓN 91. *Si f es una función medible no negativa y E es un conjunto Lebesgue medible, se define:*

$$\int_E f dm = \int_{\mathbb{R}^2} I_E f dm$$

TEOREMA 92 (**Teorema de la convergencia monótona**). *Sea f_n una sucesión no decreciente de funciones medibles no negativas, entonces:*

$$\int_{\mathbb{R}} \lim_{n \rightsquigarrow \infty} f_n dm = \lim_{n \rightsquigarrow \infty} \int_{\mathbb{R}} f_n dm$$

PROPOSICIÓN 93. *Sean f y g dos funciones medibles no negativas, entonces:*

(i) $\int_{\mathbb{R}} (af + bg)\, dm = a \int_{\mathbb{R}} f dm + b \int_{\mathbb{R}} g dm$ *para cualesquiera números reales a y b no negativos.*
(ii) *Si $f \leq g$, entonces $\int_{\mathbb{R}} f dm \leq \int_{\mathbb{R}} g dm$.*

DEFINICIÓN 94. *Se dice que una función medible f es integrable sobre un conjunto Lebesgue medible E si $\int_E |f|\, dm < \infty$.*

PROPOSICIÓN 95. *Una función medible f es integrable sobre un conjunto Lebesgue medible E si y sólo si f^+ y f^- son integrables sobre E.*

DEFINICIÓN 96. *Si f una función medible e integrable sobre un conjunto Lebesgue medible E, se define su integral sobre E, $\int_E f dm$, de la siguiente manera:*

$$\int_E f dm = \int_E f^+ dm - \int_E f^- dm$$

■

PROPOSICIÓN 97. *Sean f y g dos funciones medibles e integrables sobre un conjunto Lebesgue medible E, entonces:*

(i) *Para cualesquiera números reales a y b, la función $af + bg$ es integrable sobre E y:*
$$\int_E (af + bg)\, dm = a \int_E f dm + b \int_E g dm$$
(ii) *Si $f \leq g$ sobre E, entonces $\int_E f dm \leq \int_E g dm$.*
(iii) $\left| \int_E f dm \right| \leq \int_E |f|\, dm$

COROLARIO 98 (**Teorema de la convergencia dominada**). *Sea g una función no negativa integrable sobre un conjunto Lebesgue medible E y $\{f_n\}_{n \in N}$ una sucesión de funciones medibles tales que $|f_n| \leq g$ y $\lim_{n \rightsquigarrow \infty} f_n$ existe c.s., entonces:*

$$\int_E \lim_{n \rightsquigarrow \infty} f_n dm = \lim_{n \rightsquigarrow \infty} \int_E f_n dm$$

PROPOSICIÓN 99. *Sea* $f : [a,b] \times [c,d] \mapsto \mathbb{R}$ *una función Riemann integrable no negativa, entonces* f *es medible y la integral de Lebesgue de* f *coincide con su integral de Riemann.*

PROPOSICIÓN 100 (**Teorema de Tonelli**). *Sea* $f : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$ *una función boreliana no negativa. Para cada* $x, y \in \mathbb{R}$, *definamos las funciones* $f_x : \mathbb{R} \mapsto \overline{\mathbb{R}}$ *y* $f_y : \mathbb{R} \mapsto \overline{\mathbb{R}}$ *de la siguiente manera:* $f_x(z) = f(x,z)$ *y* $f_y(z) = f(z,y)$, *respectivamente, entonces:*

(i) f_x *es boreliana para cualquier* $x \in \mathbb{R}$.
(ii) f_y *es boreliana para cualquier* $y \in \mathbb{R}$.
(iii) *Las funciones* $x \mapsto \int_{\mathbb{R}} f_x(y)dy$ *y* $x \mapsto \int_{\mathbb{R}} f_y(x)dx$ *son borelianas.*
(iv) $\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f_x(y)dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f_y(x)dx \right) dy$

PROPOSICIÓN 101 (**Teorema de Fubini**). *Sea* $f : \mathbb{R}^2 \mapsto \overline{\mathbb{R}}$ *una función medible e integrable. Para cada* $x, y \in \mathbb{R}$, *definamos las funciones* $f_x : \mathbb{R} \mapsto \overline{\mathbb{R}}$ *y* $f_y : \mathbb{R} \mapsto \overline{\mathbb{R}}$ *de la siguiente manera:* $f_x(z) = f(x,z)$ *y* $f_y(z) = f(z,y)$, *respectivamente, entonces:*

(i) f_x *es boreliana para cualquier* $x \in \mathbb{R}$ *y* $\{x \in \mathbb{R} : f_x \textit{ no es integrable}\}$ *tiene medida cero.*
(ii) f_y *es boreliana para cualquier* $y \in \mathbb{R}$ *y* $\{y \in \mathbb{R} : f_y \textit{ no es integrable}\}$ *tiene medida cero.*
(iii) *Las funciones* $x \mapsto \int_{\mathbb{R}} f_x(y)dy$ *y* $y \mapsto \int_{\mathbb{R}} f_y(x)dx$ *son borelianas e integrables.*
(iv) $\left(\int_{\mathbb{R}} f_x(y)dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f_y(x)dx \right) dy = \int_{\mathbb{R}^2} f dm$

Respuestas a los ejercicios

CAPÍTULO 6

6.1. $P\left[|X-1|>2\right]=1-F_X(3)+F_X(-1)$

6.2. *a)* $\frac{3}{4}$; *b)* $\frac{1}{2}$; *c)* $\frac{1}{4}$; *d)* 0

6.4. *a)* $c=3+\frac{1}{2}\sqrt{2}$; *b)* $F_X(x)=\begin{cases} 0 & si\ x<-1 \\ \frac{1}{2}\left(x+1\right)^2 & si\ x\in[-1,0) \\ \frac{1}{2} & si\ x\in[0,3) \\ \frac{1}{2}+\left(x-3\right)^2 & si\ x\in[3,c) \\ 1 & si\ x\geq c \end{cases}$

6.5. *No*

6.6. *Mayor que* 0.6019

6.7. $f(x)=\begin{cases} \frac{x-1}{n}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-2}{n}\right) & si\ x\in\{2,\ldots,n+1\} \\ 0 & en\ otro\ caso \end{cases}$

CAPÍTULO 7

7.1. $\frac{135}{1024}$

7.2. $P\left[\textit{2 hombres y 2 mujeres}\right]=\frac{3}{8}$; $P\left[\textit{3 hijos de un sexo y 1 del otro}\right]=\frac{1}{2}$

7.3. $P\left[\textit{obtener por lo menos 1 seis al lanzar 6 veces un dado}\right]=0.6651$

$P\left[\textit{obtener por lo menos 2 seises al lanzar 12 veces un dado}\right]=0.61867$

$P\left[\textit{obtener por lo menos 3 seises al lanzar 18 veces un dado}\right]=0.59735$

7.4. 0.94208

7.5. 375

7.6. $\frac{k}{n}$

7.8. *Los* $\binom{n}{k}$ *posibles arreglos de* k *éxitos y* $n-k$ *fracasos tienen la misma probabilidad.*

7.9. *El jugador A pues la probabilidad que tiene de ganar el juego es igual a* 0.5706.

7.10. $f_X(x) = \begin{cases} \binom{2N-x}{N} \left(\frac{1}{2}\right)^{2N-x} & si\ x \in \{0, \ldots, N\} \\ 0 & en\ otro\ caso \end{cases}$

7.11. $P[exactamente\ k\ mujeres] = \begin{cases} \frac{2\alpha}{2-p} \left(\frac{p}{2-p}\right)^k & si\ k \in \{1, 2, \ldots\} \\ 1 - \frac{\alpha p}{(1-p)(2-p)} & si\ k = 0 \end{cases}$

7.12. $P(X \geq x) = (1-p)^x$

7.14. $\frac{n}{2^{n-1}} \left(1 - \frac{n}{2^{n-1}}\right)^{k-1}$

7.15. $f_X(x) = \begin{cases} p_1 & si\ x = 0 \\ (1-p_1)^x (1-p_2)^{x-1} \left[p_2 + p_1(1-p_2)\right] & si\ x \in \{1, 2, \ldots\} \\ 0 & en\ otro\ caso \end{cases}$

$p = p_1 + p_2(1-p_1)$

7.17. $f_X(x) = \begin{cases} \frac{1}{x!} \left(0.6e^{-2}2^x + 0.4e^{-3}3^x\right) & si\ x \in \{0, 1, \ldots\} \\ 0 & en\ otro\ caso \end{cases}$

7.18. $\frac{17}{2} e^{-3}$

7.19. $p > 1 - 2^{-10^{-11}}$

7.21. $\lambda = k$

7.22. 0.045822

7.23. $1 - \frac{5}{2} e^{-1}$

7.24. $f_X(x) = \begin{cases} \frac{x-1}{\binom{N}{2}} & si\ x \in \{2, \ldots, N\} \\ 0 & en\ otro\ caso \end{cases}$

7.25. $f_X(x) = \begin{cases} \frac{3(N-x)(N-x-1)}{N(N-1)(N-2)} & si\ x \in \{1, \ldots, N-2\} \\ 0 & en\ otro\ caso \end{cases}$

$f_Y(y) = \begin{cases} \frac{6(y-1)(N-y)}{N(N-1)(N-2)} & si\ y \in \{2, \ldots, N-1\} \\ 0 & en\ otro\ caso \end{cases}$

$f_Z(z) = \begin{cases} \frac{3(z-1)(z-2)}{N(N-1)(N-2)} & si\ z \in \{3, \ldots, N\} \\ 0 & en\ otro\ caso \end{cases}$

7.26. 0.77778

7.28. $f_X(x) = \begin{cases} \frac{4-x}{10} & si\ x \in \{0, 1, 2, 3\} \\ 0 & en\ otro\ caso \end{cases}$

7.29. $c = 6$; $P\left[X\ es\ par\right] = \frac{3}{4}$

7.31. *a)* $\frac{3}{4}$; *b)* $\frac{1}{10}$; *c)* $\frac{1}{25}$; *d)* $\frac{3}{50}$

7.32. $f_X(x) = \begin{cases} \frac{x+1}{100} & si\ x \in \{0,\dots,9\} \\ \frac{19-x}{100} & si\ x \in \{10,\dots,18\} \\ 0 & en\ otro\ caso \end{cases}$

7.33. $P[X < Y] = 1 - \frac{N(N+1)}{2MN}$; $P[X = Y] = \frac{1}{M}$

7.34. $f_Y(y) = \begin{cases} \frac{M}{N} & si\ y = M \\ \frac{1}{N} & si\ y \in \{M+1,\dots,N\} \\ 0 & en\ otro\ caso \end{cases}$

7.35. $f_Y(y) = \begin{cases} \frac{N-M+1}{N} & si\ y = M \\ \frac{1}{N} & si\ y \in \{1,\dots,M-1\} \\ 0 & en\ otro\ caso \end{cases}$

7.36. $f_Z(z) = \begin{cases} \frac{2z-1}{N^2} & si\ z \in \{1,\dots,N\} \\ 0 & en\ otro\ caso \end{cases}$

7.37. *a)* $P[X < Y] = 1 - e^{-\lambda p}$; $P[X > Y] = (1-p)e^{-\lambda p}$

b) $P[X > Y] > P[X < Y]$ *cuando* $\lambda < \frac{1}{p}\ln(2-p)$.

7.38. *a)* $P[X < Y] = \frac{1-p}{2-p}$; $P[X = Y] = \frac{p}{2-p}$

b) $f_Z(z) = \begin{cases} p(2-p)(1-p)^{2z} & si\ z \in \{0,1,\dots\} \\ 0 & en\ otro\ caso \end{cases}$

CAPÍTULO 8

8.1. T *tiene distribución uniforme en el intervalo* $[0,15]$.

8.2. $f_Y(y) = \begin{cases} 2(1-y) & si\ y \in (0,1) \\ 0 & en\ otro\ caso \end{cases}$

8.3. *a)* Y *tiene distribución uniforme en el intervalo* $(0,1)$.

b) $f_Z(z) = \begin{cases} \frac{1}{2\sqrt{z}} & si\ 0 < z < 1 \\ 0 & en\ otro\ caso \end{cases}$

8.4. $f_Y(y) = \begin{cases} \frac{2}{\pi\sqrt{(1-y^2)}} & si\ y \in (0,1) \\ 0 & en\ otro\ caso \end{cases}$

8.5. $\frac{T-t}{T-pt}p$

8.6. 0.682689

8.7. 0.354032

8.8. 0.193939

8.9. 0.10618

8.10. $f_X(x) = \begin{cases} \frac{1}{\int_b^\infty \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}dy} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} & \textit{si } x \geq b \\ 0 & \textit{en otro caso} \end{cases}$

8.11. *Utilizando el teorema de de Moivre-Laplace, se tiene* $P[X = 310] \approx 0.0216969$*, mientras que el valor exacto es* 0.0215234*.*

$P[280 < X < 310] \approx 0.654736$

8.12. 0.998611

8.13. 1

8.14. 0.36529

8.15. 508

8.16. 0.220718

8.17. 9604

8.18. 0.0312832

8.19. 1.52

8.20. 2.3685×10^{-7} *años*

8.21. 2 *segundos.*

8.23. $f_Y(t) = \begin{cases} 30t\,(1+t)^2\, e^{-3t}\left[1-(1+t)\,e^{-t}\right]^2 & \textit{si } t > 0 \\ 0 & \textit{en otro caso} \end{cases}$

8.24. 0.761897

8.25. $1 + e^{-5\lambda}(5\lambda + 1) - e^{-\lambda}(\lambda + 1)$

8.26. Y *tiene distribución gama con parámetros* α *y* $\frac{\lambda}{c}$*.*

8.27. *a)* $\left(\frac{2}{3}\right)^4$*; b)* $\left(\frac{1}{3}\right)^4$

8.28. $f_Y(y) = \begin{cases} 2(1-y) & \textit{si } y \in (0,1) \\ 0 & \textit{en otro caso} \end{cases}$

8.29. $f_X(x) = \begin{cases} \frac{x}{\sqrt{x^2-3}} & \textit{si } x \in \left(\sqrt{3}, 2\right) \\ 0 & \textit{en otro caso} \end{cases}$

8.30. $f_X(x) = \begin{cases} \frac{1}{2}\pi x & \textit{si } 0 \leq x < 1 \\ 2x arcsen \frac{1}{x} - \frac{1}{2}\pi x & \textit{si } 1 < x \leq \sqrt{2} \\ 0 & \textit{en otro caso} \end{cases}$

8.31. $f_Z(z) = \begin{cases} z & \textit{si } 0 \leq z \leq 1 \\ 2 - z & \textit{si } 1 < z \leq 2 \\ 0 & \textit{en otro caso} \end{cases}$

8.32. $f_X(x) = \begin{cases} 2\sqrt{2}\left(1 - x\sqrt{2}\right) & \textit{si } x \in \left(0, \frac{1}{2}\sqrt{2}\right) \\ 0 & \textit{en otro caso} \end{cases}$

8.33. $f_X(x) = \begin{cases} \frac{1}{2}(2 - x) & \textit{si } 0 \leq x \leq 2 \\ 0 & \textit{en otro caso} \end{cases}$

8.34. $d = 1 - \frac{1}{2}\sqrt{2}$

8.35. $f_Y(y) = \frac{1}{|a|} f(\frac{y-b}{a})$

Para el caso en que X tiene distribución uniforme en el intervalo $(0,1)$, si $a > 0$, Y tiene distribución uniforme en el intervalo $(b, a+b)$, mientras que si $a < 0$, Y tiene distribución uniforme en el intervalo $(a+b, b)$.

8.36. $f_Y(y) = \begin{cases} 2yf(y^2) & \textit{si } y > 0 \\ 0 & \textit{en otro caso} \end{cases}$

Para el caso en que X tiene distribución gama con parámetros α y λ, se tiene:

$f_Y(y) = \begin{cases} \frac{2\lambda^\alpha}{\Gamma(\alpha)} y^{2\alpha-1} e^{-\lambda y^2} & \textit{si } y > 0 \\ 0 & \textit{en otro caso} \end{cases}$

8.37. $f_Y(y) = \begin{cases} f(y) + f(-y) & \textit{si } y > 0 \\ 0 & \textit{en otro caso} \end{cases}$

a) $f_Y(y) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \left[e^{-(y-\mu)^2/2\sigma^2} + e^{-(y+\mu)^2/2\sigma^2} \right] & \textit{si } y > 0 \\ 0 & \textit{en otro caso} \end{cases}$

b) $f_Y(y) = \begin{cases} \frac{2}{3} & \textit{si } y \in (0,1) \\ \frac{1}{3} & \textit{si } y \in [1,2) \\ 0 & \textit{en otro caso} \end{cases}$

8.38. $f_Y(y) = \begin{cases} \frac{1}{y} f(\ln y) & \textit{si } y > 0 \\ 0 & \textit{en otro caso} \end{cases}$

a) $f_Y(y) = \begin{cases} \frac{1}{\sigma y\sqrt{2\pi}} e^{-(\ln y - \mu)^2/2\sigma^2} & \textit{si } y > 0 \\ 0 & \textit{en otro caso} \end{cases}$

b) $f_Y(y) = \begin{cases} \frac{1}{y^{\lambda+1}\Gamma(\alpha)}\lambda^{\alpha}(\ln y)^{\alpha-1} & si\ y > 1 \\ 0 & en\ otro\ caso \end{cases}$

8.39. $f_Y(y) = \begin{cases} \frac{3y-4\sqrt{y}+5}{8\sqrt{y}} & si\ 0 < y < 1 \\ 0 & en\ otro\ caso \end{cases}$

CAPÍTULO 9

9.3. $\frac{2N+1}{3}$

9.4. $e-1$

9.5. $-\frac{17}{216}$

9.6. *a)* $\frac{n}{6} - n\left(\frac{5}{6}\right)^n$; *b)* $n = 10$

9.7. $\frac{n(N+1)}{n+1}$

9.8. $N - \frac{1}{N^n}\sum_{k=1}^{N-1} k^n$

9.9. $E(Y) = \frac{1-p}{p}\left[1-(1-p)^M\right]$; $E[Z] = M + \frac{(1-p)^{M+1}}{p}$

9.10. *a)* 5.5; *b)* 3.4

9.11. 4

9.12. N

9.13. *a)* n; *b)* $\frac{n+1}{2}$

9.14. $\frac{2+p(1-p)}{1-p(1-p)}$

9.15. $N+1$

9.16. $\frac{1}{n}2^{n-1}$

9.17. 2

9.18. 3

9.19. 15

9.20. 22.22

9.21. 3

9.22. *En los dos casos el número esperado de cajones que se abrirán hasta encontrar el documento es igual a* 15.5.

9.23. *En el caso (a), el número esperado de cajones que se abrirán hasta encontrar el documento es igual a* $\frac{1}{2}\frac{82-71p}{2-p}$*, mientras que en el caso (b) es igual a* $\frac{3}{2}\frac{14+3p}{2-p}$*, de manera que la mejor estrategia es la indicada en (b).*

9.24. *a)* 0.3012*; b)* 99.32

9.25. 0.761897

9.26. 0.1844

9.27. $\frac{1}{(n+1)p}\left(1-(1-p)^{n+1}\right)$

9.28. $\sqrt{2\pi e}\,[1-F(1)]=0.65568$

9.29. $\frac{e^{3/2}}{\sqrt{3}}$

9.31. $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}$

9.32. 35

9.33. 1

9.34. $\frac{n+1}{r+1}$

9.35. $\frac{s(n+1)}{r+1}$

9.36. $N\left[1+\frac{1}{m}-(1-p)^m\right]$

9.37. $365\left[1-\left(\frac{364}{365}\right)^N\right]$

9.38. $N\left(1+\frac{1}{2}+\cdots+\frac{1}{N}\right)$

9.40. *a)* $\frac{1}{2}(a+b)$*; b)* μ*; c)* $\frac{1}{\lambda}\ln 2$

9.42. *A una distancia igual a* $\frac{1}{\lambda}\ln 2$ *del origen.*

9.43. $E(X)=\frac{1}{2}N$; $Var(X)=\frac{N(N+2)}{12}$

9.44. $E(X)=\frac{1}{2}(a+b)+\frac{1}{2n}(b-a)$; $Var(X)=\frac{1}{12}(b-a)^2-\frac{1}{12n^2}(b-a)^2$

9.45. $E\left[|X+1|\right]=\frac{5}{6}$; $Var\left[|X+1|\right]=\frac{11}{36}$

9.46. $E\left[Y\right]=\frac{28}{15}$; $Var(Y)=\frac{866}{225}$

9.47. 299.67

9.48. 2

9.49. $4Var(X)+9Var(Y)$

9.50. $E\left[X\right]=\frac{100}{19}$; $Var(X)=2.6397$

9.51. $Var(X) = \frac{nrs(r+s-n)}{(r+s)^2(r+s-1)}$

9.52. $E[X^n] = \frac{\Gamma(\alpha+n)}{\lambda^n \Gamma(\alpha)}$

9.53. *b)* $E[(X-\mu)^n] = \begin{cases} 0 & si\ n\ es\ impar \\ 1 \cdot 3 \cdot 5 \cdots (n-1)\sigma^n & si\ n\ es\ par \end{cases}$

9.55. *La desigualdad de Chebyshev establece que a lo más se rechazan* 1 *de cada* 10 *monedas balanceadas.*

El teorema de de Moivre Laplace establece que se rechazan, aproximadamente, 16 *de cada* 10, 000 *monedas balanceadas.*

9.56. 0.31731

9.57. *a)* $\frac{15}{17}$*; b) es mayor o igual a* $\frac{3}{4}$*; c) por lo menos* 10

9.58. *a)* $P[12 \leq X \leq 18] = 0.79951$*; b)* $P[12 \leq X \leq 18] \geq \frac{1}{6}$

9.59. *a)* $n \geq 9604$*; b* $n \geq 50000$

9.60. $\Phi_X(t) = te^{\lambda(t-1)}$

9.61. $\Phi_X(t) = \frac{2^N t + t^{N+1} - t^{N+2}}{2^N(2-t)}$

9.62. $\Phi_X(t) = \begin{cases} \frac{t}{N}\frac{1-t^N}{1-t} & si\ t \neq 1 \\ 1 & si\ t = 1 \end{cases}$

$E[X] = \frac{N+1}{2}$

9.63. $\Phi_X(t) = \frac{2t^3}{N+2-Nt}$; $E[X] = 3 + \frac{1}{2}N$

9.64. $\Phi_X(t) = (tp + 1 - p)^n$; $E[X] = np$; $Var(X) = np(1-p)$

9.65. $\frac{35}{2}$

9.66. $\frac{3055}{36}$

9.67. $f_{X+Y}(1) = \frac{3}{90}$; $f_{X+Y}(2) = \frac{2}{90}$; $f_{X+Y}(3) = \frac{19}{90}$; $f_{X+Y}(4) = \frac{13}{90}$; $f_{X+Y}(5) = \frac{31}{90}$; $f_{X+Y}(6) = \frac{10}{90}$; $f_{X+Y}(7) = \frac{12}{90}$

$f_{X+Y}(z) = 0$ *para* $z \notin \{1, 2, 3, 4, 5, 6, 7\}$.

9.69. $M_X(t) = \frac{1}{2(1+t)} + \frac{1}{2(1-t)}$ *para* $x \in (-1, 1)$.

9.71. *a)* $M_U(t) = \begin{cases} \frac{1}{t^2}(e^t - 1)^2 & si\ t \neq 0 \\ 1 & si\ t = 0 \end{cases}$, *b)* $M_V(t) = \begin{cases} \frac{1}{t^2}e^{-t}(e^t - 1)^2 & si\ t \neq 0 \\ 1 & si\ t = 0 \end{cases}$

9.73. 278

Tabla de la distribución normal estándar

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{1}{2}y^2} dy$$

$0 \quad z$

z	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3079	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4430	.4441
1.6	.4452	.4463	.4474	.4485	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4700	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4762	.4767
2.0	.4773	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4865	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4980	.4980	.4981
2.9	.4981	.4982	.4983	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									