UN PRIMER CURSO DE PROBABILIDAD

L. A. Rincón
Departamento de Matemáticas
Facultad de Ciencias, UNAM
Circuito Exterior de CU
04510 México DF
MÉXICO
lars@fciencias.unam.mx

Junio-Noviembre 2001

Contenido

1 Preliminares **3**
- 1.1 Historia 4
- 1.2 Conjuntos 7

2 Conceptos Básicos **11**
- 2.1 Experimentos Aleatorios, Espacios Muestrales y Eventos 11
- 2.2 Probabilidad Clásica y Frecuentista 14
- 2.3 Análisis Combinatorio 16
- 2.4 Probabilidad Geométrica 23
- 2.5 Axiomatización de la Probabilidad 24
- 2.6 Probabilidad Condicional 31
- 2.7 Independencia de Eventos 38

3 Variables Aleatorias **43**
- 3.1 Variables Aleatorias 43
- 3.2 Funciones de Densidad y de Distribución 46
- 3.3 Distribuciones Discretas 53
- 3.4 Distribuciones Continuas 60
- 3.5 La Distribución Normal 65

4 Esperanza, Varianza, Momentos **69**
- 4.1 Esperanza 69
- 4.2 Esperanza de Funciones de Variables Aleatorias . . . 76

4.3 Varianza 78
4.4 Momentos 82
4.5 La Función Generadora de Momentos 84

5 Teoremas Límite **88**
5.1 Vectores Aleatorios 88
5.2 La Desigualdad de Chebyshev 93
5.3 La Ley de los Grandes Números 96
5.4 El Teorema de De Moivre-Laplace 98
5.5 El Teorema del Límite Central 100

Capítulo 1

Preliminares

Empezaremos nuestro curso de probabilidad definiendo lo que entenderemos por el término ”teoría de la probabilidad”, revisaremos brevemente cuál ha sido el desarrollo histórico de la probabilidad y finalmente recordaremos cierta terminología de la teoría de conjuntos que nos será de utilidad.

¿ Qué es la Teoría de la Probabilidad ?

La teoría de la probabilidad es el estudio matemático de fenómenos aleatorios, esto es, de fenómenos gobernados por el azar.

Intentar definir lo que es el azar es un tanto complicado de modo que dejaremos nuestra definición a un nivel intuitivo. Más adelante trataremos de explicar con mayor precisión el tipo de fenómenos aleatorios que estudiaremos.

1.1 Historia

Presentamos a continuación una breve exposición del desarrollo histórico de la probabilidad.

Antecedentes históricos

Mencionaremos primeramente a dos personajes y sus trabajos, relacionados con el estudio incipiente de fenómenos azarosos.

Gerolamo Cardano (1510-1576) originario de Milán, publica un pequeño manual titulado “liber de ludo aleae”. En este ensayo el autor describe el espacio muestral del experimento de lanzar dos y tres dados.

Galileo Galilei (1564-1632) plantea y discute la “Teoría de los errores” que es una de las primeras aplicaciones de la probabilidad a la fisica. Galileo llega a la conclusión de que los errores aleatorios son inevitables en las observaciones instrumentales. También concluye que los errores pequeños son más frecuentes que los mayores. En su obra “Considerazione sopra il gluco del dadi” estudia las posibles combinaciones que pueden obtenerse al lanzar tres dados.

Origen de la teoría clásica de la probabilidad

Antonio de Gombaud fue un noble francés aficionado al juego. Era conocido popularmente como el caballero De Mere y planteó a Blaise Pascal (1623-1662) el siguiente problema:

De Mere y su compañero de juego escogen cada uno un número del 1 al 6, y lanzan sucesivamente un dado apostando 32 doblones de oro cada uno, a que el número escogido aparece en tres ocasiones antes que el del contrario. El ganador obtiene los 64 doblones de oro. Si el dado es honesto, cada uno tiene la misma probabilidad de ganar, de modo que el juego es justo. Pero ellos se plantean el dilema de qué hacer en caso de que uno de los números tenga dos apariciones y el otro una aparición, y por alguna razón tengan que suspender el juego, ¿ Cómo tendrían que repartirse los 64 doblones si se llegase a tal situación ?

B. Pascal decide consultar a Pierre de Fermat (1601-1665) acerca del problema planteado por De Mere. Pascal y Fermat mantienen conrrespondencia acerca del problema. Sobre esta correspondencia sobreviven tres cartas fechadas en el año 1654. Este episodio constituye el inicio de la teoría clásica de la probabilidad como se conoce hoy en dia.

Desarrollo

En ésa época no existía la teoría de la probabilidad como la conocemos hoy en dia, simplemente se tenían algunos métodos de solución a ciertos problemas relacionados con juegos de azar. Los procedimientos de solución de estos problemas usaban aritmética elemental, cálculo combinatorio y en ese entonces se establecieron los conceptos de permutación, combinación y coeficiente binomial entre otros.

La teoría elemental de la probabilidad fue enriquecida por los trabajos de: Huyghens, James y Jacob Bernoulli, A. deMoivre, A. M. Legendre, J.L. Lagrange, Laplace, Gauss, Poisson, etc.

En 1812, P. S. Laplace publicó el libro "Theorie analytique des probabilités" en el cual se introducen nuevos conceptos, se sistematizan y se extienden resultados importantes obtenidos hasta ese entonces.

De 1850 a 1920 el desarrollo de la teoría de la probabilidad está asociado fundamentalmente a los trabajos de los matemáticos rusos:

Chebyshev, Markov y Lyapunov entre otros.

El rápido desarrollo de la teoría es provocado tanto por preguntas sin resolver de la misma teoría como por la amplia gama de sus aplicaciones.

En la primera década del siglo XX Emile Borel introdujo algunas ideas que conectaron la teoría de la probabilidad con cuestiones teóricas de medida usando funciones de variable real.

En la década de 1920 se desarrollaron las ideas de Borel y resultaron muy fructíferas para el desarrollo de la teoría. Entre los matemáticos involucrados en este desarrollo están: Khintchine, Kolmogorov, Eugene Slutsky, Paul Lévy, Anton Lomnitsky, Lindeberg, Bernstein y William Feller entre otros. Se lograron resolver de manera definitiva algunos problemas clásicos formulados en la época de Chebyshev.

El campo de acción de la teoría de la probabilidad fué expandido considerablemente con las nociones de teoría de la medida y del análisis funcional.

En la década de 1930 se crea la teoría de los procesos estocásticos. En 1933 A. N. Kolmogorov axiomatiza la probabilidad. Esto es, no se establece la forma de calcular probabilidades sino que se proponen las propiedades que ésta debe saisfacer. El trabajo de muchos matemáticos y en particular de Emile Borel y Norbert Wiener sirvieron de antecedente para la axiomatización presentada por Kolmogorov. Desde entonces la teoría se ha desarrollado muy rápidamente y sus métodos has servido para estudiar y comprender fenómenos de la fisica, la quimica, y en general fenómenos en los que por alguna razón sea necesario considerar el azar.

Se tienen por ejemplo del conceptos de movimiento Browniano y el de ecuación diferencial estocástica de amplia aplicación.

El lector interesado en leer con mayor detalle la historia de la probabilidad puede encontrar mayor información en [12] y en [16].

1.2 Conjuntos

En esta pequeña sección recordaremos brevemente algunas operaciones entre conjuntos asi como algunas propiedades que nos serán de utilidad en el estudio de la probabilidad.

Suponemos dado un conjunto universal de objetos cualesquiera, que denotaremos por la letra griega Ω (Omega). Y usaremos la letra minúscula ω para denotar un elemento cualquiera del conjunto Ω.

Operaciones Básicas

Si A y B son dos subconjuntos cualesquiera de Ω, recordemos las operaciones básicas de unión, intersección, diferencia y complemento, respectivamente:

$$\begin{aligned} A \cup B &= \{\omega \in \Omega : \omega \in A \text{ ó } \omega \in B\}, \\ A \cap B &= \{\omega \in \Omega : \omega \in A \text{ y } \omega \in B\}, \\ A - B &= \{\omega \in \Omega : \omega \in A \text{ y } \omega \notin B\}, \\ A^c &= \{\omega \in \Omega : \omega \notin A\} \end{aligned}$$

Definimos también la operación **diferencia simétrica** entre dos conjuntos A y B, denotada por $A \triangle B$, como sigue

$$A \triangle B = (A - B) \cup (B - A).$$

Definición. *Decimos que dos conjuntos A y B son* **ajenos** *(o disjuntos) si y solo si*

$$A \cap B = \emptyset.$$

Analogamente decimos que n conjuntos $A_1, A_2, \ldots, A_n$ son **ajenos** *(o mutuamente ajenos) si*

$$A_1 \cap A_2 \cap \cdots \cap A_n = \emptyset.$$

Interpretación

Cuando los conjuntos se expresan en palabras, la operación unión, $A \cup B$, se lee "A o B" y la intersección, $A \cap B$, se lee "A y B". El complemento A^c se interpreta como aquellos elementos que estan fuera de A. Cuando A y B son ajenos, esto es, $A \cap B = \emptyset$, no hay ningún elemento común entre A y B.

El conjunto vacío y el conjunto universal satisfacen las siguientes propiedades elementales

$$\begin{array}{rl} A \cup \emptyset = A, & \quad A \cap \emptyset = \emptyset, \\ A \cup \Omega = \Omega, & \quad A \cap \Omega = A, \\ A \cup A^c = \Omega, & \quad A \cap A^c = \emptyset. \end{array}$$

Asociatividad/Distributividad

Las operaciones unión e intersección son asociativas, esto es,

$$\begin{aligned} A \cup (B \cup C) &= (A \cup B) \cup C, \\ A \cap (B \cap C) &= (A \cap B) \cap C. \end{aligned}$$

Y recordemos también las siguientes propiedades distributivas que nos serán de utilidad

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C), \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C). \end{aligned}$$

Leyes de De Morgan

Finalmente recordemos las leyes de De Morgan que nos serán de mucha utilidad en el cálculo de probabilidades.

$$
\begin{aligned}
(A \cup B)^c &= A^c \cap B^c, \\
(A \cap B)^c &= A^c \cup B^c.
\end{aligned}
$$

Estas igualdades no solo son ciertas para dos eventos A y B como lo hemos expresado arriba, sino que siguen siendo válidas para cuando tenemos un número finito cualquiera de eventos. E incluso para uniones e intersecciones infinitas arbitrarias.

EJERCICIOS TEÓRICOS

1. Sean A, B y C tres subconjuntos arbitrarios de Ω. Exprese $A \cup B \cup C$ como la unión de tres conjuntos ajenos. Utilice un diagrama de Venn para ilustrar su solución. Observe que no existe una única solución.

2. Use las propiedades de las operaciones entre conjuntos para demostrar rigurosamente las siguientes igualdades. En cada caso dibuje un diagrama de Venn para ilustrar la validez de cada igualdad.

   (a) $A = (A \cap B) \cup (A \cap B^c)$.

   (b) $A^c - B^c = B - A$.

   (c) $A \cap B^c = A - (A \cap B)$.

   (d) $A \cup B = B \cup (A \cap B^c) = A \cup (B \cap A^c)$.

   (e) $A - (B \cap C) = (A - B) \cup (A - C)$.

3. Demuestre formalmente las siguientes proposiciones. En cada caso dibuje un diagrama de Venn para ilustrar la validez de cada proposición.

(a) $A \cap B \subset A \subset A \cup B$.

(b) Si $A \cap B = \emptyset$ entonces $A \subset B^c$.

(c) Si $A \subset B$ entonces $B^c \subset A^c$.

(d) Si $A \cap B = \emptyset$ entonces $A \cup B^c = B^c$.

(e) Si $A \subset B$ entonces $A \cup (B - A) = B$.

4. Demuestre que $A \triangle B = (A \cup B) - (A \cap B)$.

5. ¿ Es la diferencia simétrica una operación conmutativa, i.e $A \triangle B = B \triangle A$?
   ¿ Es asociativa, i.e. $(A \triangle B) \triangle C = A \triangle (B \triangle C)$?
   ¿ Tiene elemento neutro, i.e. existe un conjunto N tal que $A \triangle N = N \triangle A = A$?

6. Demuestre que $A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$ y proporcione un contraejemplo para demostrar que en general la igualdad $A \cup (B \triangle C) = (A \cup B) \triangle (A \cap C)$ es falsa.

EJERCICIOS DE APLICACIÓN

1. El 40% de las mujeres fuman mientras que el 30% de los hombres son fumadores. Consideremos una población de igual número de mujeres que de hombres. ¿ Qué porcentaje total de la población son fumadores ?

Capítulo 2

Conceptos Básicos

2.1 Experimentos Aleatorios, Espacios Muestrales y Eventos

Esta pequeña sección contiene algunos conceptos elementales de la teoría de la probabilidad. Hemos mencionado antes que la teoría de la probabilidad se encarga de la modelación matemática de los experimentos aleatorios. Explicaremos ahora con más precisión este tipo de experimentos.

¿ Qué es un Experimento Aleatorio ?

Estos son experimentos que no siempre arrojan los mismos resultados aún cuando se les repita bajo las mismas condiciones iniciales. Los resultados de los mismos dependen de alguna característica azarosa que no es posible controlar o predecir. Ejemplos sencillos de experimentos aleatorios son: el lanzamiento de una moneda o de un dado, un juego de lotería, el registro del tiempo de espera para obtener algún servicio, el precio en la cotización de un bien económico que fluctúa de acuerdo a la oferta y la demanda, la observación diaria del comportamiento de cierto índice económico, etc.

¿ Qué es un Espacio Muestral ?

Asociado a todo experimento aleatorio tendremos siempre un conjunto que constará de todos los posible resultados del experimento aleatorio. A este conjunto le llamaremos el **espacio muestral** y lo denotaremos por la letra griega Ω (Omega). Este conjunto representa nuestro universo o totalidad de resultados del experimento aleatorio. Por ejemplo, si el experimento aleatorio consiste en lanzar un dado entonces claramente $\Omega = \{1, 2, 3, 4, 5, 6\}$. A un elemento cualquiera de este conjunto se le llama **punto muestral**, y lo denotaremos por la letra omega minúscula ω.

Decimos que un espacio muestral es **discreto** si consta de un número finito o infinito numerable de elementos. Si el espacio muestral consta de un número infinito no numerable de eventos entonces decimos que es **continuo**.

¿ Qué es un Evento ?

Preliminarmente definiremos un **evento** como cualquier subconjunto del espacio muestral. Para denotar a los eventos usaremos las primeras letras del alfabeto en mayúsculas, esto es, A, B, C, etc. Posteriormente modificaremos ligeramente nuestra definición de evento. Un evento se llama **simple** si consta de sólo un punto muestral. Diremos que un evento es **compuesto** si consta de mas de un punto muestral. Al evento $\emptyset$ le llamaremos **evento imposible**, y al evento Ω le llamaremos **evento seguro**.

¿ Qué significa que un evento ocurra ?

Significa simplemente que al realizar el experimento aleatorio, el punto muestral que se obtiene es un elemento del evento. Por ejemplo, en el experimento de lanzar un dado, si definimos el evento $A = \{2, 4, 6\}$ y si al lanzar el dado se obtiene "2", entonces decimos que se observó el evento A. En cambio si se obtiene "5" entonces no se observó o realizó el evento A. El evento imposible $\emptyset$ nunca ocurre, y el evento Ω ocurre siempre.

EJERCICIOS

1. Describa el correspondiente espacio muestral para cada uno de los siguientes experimentos aleatorios . Diga además si el espacio muestral es finito, infinito numerable, o infinito no numerable.
   (a) Lanzar una moneda tres veces.
   (b) Escoger un número al azar dentro del intervalo $[0, 1]$.
   (c) Colocación al azar de dos bolas distinguibles en cuatro celdas numeradas.
   (d) Tiempo que un cliente de un banco tiene que esperar para ser atendido.
2. Describa tres experimentos de su interés que considere aleatorios, identifique los correspondientes espacios muestrales, clasificando estos espacios como discretos o continuos, y finalmente proporcione un ejemplo de un evento cualquiera para cada experimento sugerido.
3. Sean A, B y C tres eventos de un espacio muestral Ω. Exprese las siguientes oraciones en términos de intersecciones, uniones y complementos de los conjuntos A, B y C.
   (a) Ninguno de los eventos ocurre.

(b) Exactamente uno de los eventos ocurre.

(c) Exactamente dos de los eventos ocurren.

(d) Los tres eventos ocurren.

(e) Al menos uno de los tres eventos ocurre.

(f) A lo mas uno de ellos ocurre.

(g) Al menos dos de los eventos ocurren.

(h) A lo mas dos de ellos ocurren.

(i) Al menos tres de los eventos ocurren.

(j) A lo mas tres de ellos ocurren.

(k) Ocurre A o B, pero no ambos, y no ocurre C.

(l) Ocurre C, y solo uno de A y B.

(m) Solo A ocurre.

(n) Ocurren A y B pero no C.

4. Considere el experimento de lanzar una moneda tres veces. Proponga un espacio muestral para este experimento, considerando que

   (a) las tres monedas son distinguibles.

   (b) las tres monedas son indistinguibles.

   (c) sólo una moneda es distinguible de las otras dos.

5. Un experimento consiste en lanzar una dado hasta que se obtiene un "6". Proponga dos espacios muestrales para este experimento.

2.2 Probabilidad Clásica y Frecuentista

Sea $\mathcal{E}$ un experimento aleatorio con correspondiente espacio muestral Ω, y sea $A \subset \Omega$ un evento. Estamos interesados en calcular la probabilidad del evento A. Es decir, queremos encontrar una medida del grado de posibilidad de ocurrencia del evento A cuando se realiza una vez el experimento aleatorio.

La probabilidad de un evento A será un número que denotaremos por $P(A)$. El problema es cómo calcular esta probabilidad. Tenemos al menos dos formas de definir o asignar esta probabilidad. Las llamaremos probabilidad clásica y probabilidad frecuentista. A continuación explicaremos en qué consisten estas formas de calcular probabilidades.

¿ Qué es la probabilidad clásica ?

Esta es una definición a priori de la probabilidad. Definimos la probabilidad de un evento A como

$$P(A) = \frac{\#A}{\#\Omega},$$

en donde $\#A$ indica la cardinalidad o número de elementos del evento A. Esta definición de probabilidad sirve sólo cuando Ω consta de un número finito de elementos. Presupone además que el espacio es equiprobable, es decir, cada punto muestral tiene la misma probabilidad de ocurrir.

¿ Qué es la probabilidad frecuentista ?

Esta es una definición a posteriori de la probabilidad. Para calcular $P(A)$ procedemos como sigue. Realizamos n veces el experimento aleatorio y denotamos por n_A el número de veces que se observó la ocurrencia del evento A. Entonces definimos

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}.$$

Esta definición de probabilidad no adolece de las limitaciones mencionadas de la probabilidad clásica aunque tiene el serio inconveniente de que en la práctica no es posible realizar una infinidad de veces el experimento.

2.3 Análisis Combinatorio

Para aplicar nuestra definición de probabilidad clásica es necesario saber calcular la cardinalidad de un evento, es decir, saber exactamente cuantos elementos tiene ese evento. Cuando los elementos del evento pueden ser puestos explicitamente en una lista, es fácil contarlos. Sin embargo, a menudo no es posible listar todos los elementos de un evento y es entonces necesario conocer algunas técnicas de conteo.

Empezaremos esta sección mencionando un principio de mucha utilidad y que se encuentra en el centro de la mayoría de los cálculos de problemas de conteo.

Regla del Producto para Pares Ordenados
Sean A y B dos conjuntos con n y m elementos respectivamente. Entonces el producto Cartesiano $A \times B$ que consta de todas las parejas ordenadas (a, b) con a cualquier elemento de A y b cualquier elemento de B, tiene $n \cdot m$ elementos.

Todas las parjas ordenadas del producto $A \times B$ pueden ser representadas gráficamente mediante un "diagrama de árbol" que ilustraremos a continuación. Si $A = \{a_1, a_2, a_3\}$ y $B = \{b_1, b_2\}$ entonces (ver diagrama dibujado en clase).

Ejemplo. Si un hombre tiene 6 camisas y 7 pantalones, ¿ De cuántas formas diferentes puede vestirse ? Respuesta: De $6 \times 7 = 42$ maneras distintas.

Más generalmente, tenemos el siguiente resultado.

Regla del Producto
Sean $A_1, A_2, \ldots, A_k$ conjuntos tales que cada A_i tiene n_i elementos. Entonces el conjunto producto Cartesiano $A_1 \times A_2 \times \cdots \times A_k$ que consta de todos los vectores de la forma $(a_1, a_2, \ldots, a_k)$ con $a_i \in A_i$ tiene un total de $n_1 \cdot n_2 \cdots n_k$ elementos.

Ejemplo. Si una mujer tiene 3 sombreros, 6 blusas, 8 faldas y 10 pares de zapatos, ¿ De cuántas formas diferentes puede vestirse ? Respuesta: De $3 \times 6 \times 8 \times 10 = 1440$ maneras distintas.

¿ Qué es una permutación ?

Supongamos que tenemos un conjunto de n objetos cualesquiera y que queremos obtener muestras ordenadas de tamaño k y sin reemplazo. El orden en las muestras significa que tomamos un objeto después de otro y no todos a la vez. Esto significa entones que tenemos un primer elemento, un segundo elemento, etc. Y el no reemplazo significa que no hay repeticiones de objetos en la muestra, esto es, una vez seleccionado un objeto, no puede volver a ser seleccionado dentro de la misma muestra. El total de arreglos o muestras de la forma antes mencionada es el número

$$\frac{n!}{(n-k)!} = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1).$$

Y se le conoce como el número de permutaciones de tamaño k de un conjunto de n objetos. Respondiendo a la pregunta arriba planteada, una permutación de tamaño k es entonces un conjunto ordenado de tamaño k.

En particular, cuando $k = n$, es decir, todos los objetos son seleccionados, las diferentes permutaciones de estos n objetos es $n!$, que corresponde al total de formas en que los n objetos pueden ser ordenados.

¿ Qué es una combinación ?

Dado un conjunto de n objetos, el total de subconjuntos de tamaño k que se pueden obtener del conjunto dado es el número

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!(n-k)!}.$$

Este número representa entonces el total de muestras no ordenadas y sin reemplazo de tamaño k de un conjunto de n objetos. Se le conoce con el nombre de **coeficiente binomial**.

Una combinación de tamaño k es simplemente un subconjunto de tamaño k obtenido de un conjunto mayor. A diferencia de una permutación, aqui el orden no importa.

Ejemplo. ¿ Cuántos equipos de 3 personas se pueden formar de un grupo de 5 personas ? Respuesta: Existen $\left(\begin{array}{c} 5 \\ 3 \end{array} \right) = \frac{5!}{3!(5-3)!} = 10$ equipos diferentes.

Los coeficientes binomiales aparecen por ejemplo como coeficientes en el desarrollo de elevar a la n-ésima potencia un binomio. Este es el teorema del binomio que enunciaremos a continuación.

Teorema del binomio
Para cualesquiera a y b números reales y $n = 1, 2, 3, \ldots$

$$(a+b)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) a^k b^{n-k}$$

El lector puede realizar la demostración del teorema del binomio usando el método de inducción.

EJERCICIOS TEÓRICOS

1. Demuestre las siguientes igualdades.

   (a) $\binom{n}{k} = \frac{n-k-1}{k}\binom{n}{k-1}$.

   (b) $\binom{n}{k} = \frac{n}{n-k}\binom{n-1}{k}$.

   (c) $\binom{n-1}{k} = \frac{k+1}{n}\binom{n}{k+1}$.

2. Utilice el teorema del binomio, escogiendo adecuadamente los valores de a y b, para demostrar las siguientes identidades.

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

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

   (c) $\sum_{k=0}^{n}\binom{n}{k}(a-1)^k = a^n$.

3. Demuestre que

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

   Sugerencia: Escriba $(1+b)^{m+n} = (1+b)^m(1+b)^n$. Utilice el teorema del binomio para escribir cada uno de estos binomios en términos de sumas. Calcule el coeficiente del término b^k de ambos lados de la ecuación. Estos coeficientes deben ser iguales si es que los polinomios lo son.

4. Use el ejercicio anterior para demostrar que

$$\sum_{r=0}^{n} \binom{n}{r}^2 = \binom{2n}{n}.$$

5. Demuestre que

   (a) $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$.

   (b) Use el inciso anterior para demostrar que

$$\binom{n}{k} = \sum_{i=1}^{k+1} \binom{n-i}{k-i+1}.$$

6. Demuestre que

$$\sum_{k=0}^{n} k \binom{n}{k} = n2^{n-1}.$$

   Sugerencia: Use el teorema del binomio para $(1+b)^n$. Derive respecto a b y evalúe en $y = 1$.

EJERCICIOS DE APLICACIÓN

1. ¿ De cuántas maneras diferentes pueden clasificarse los tres primeros lugares de una carrera de 15 corredores ? Respuesta: 2730.

2. Una enciclopedia que consta de 5 volúmenes es colocada en el librero de modo aleatorio. Demuestre que la probabilidad de que los volúmenes estén colocados apropiadamente de derecha a izquierda o de izquierda a derecha es de 1/60.

3. ¿ De cuántas maneras diferentes se pueden sentar 8 personas en una mesa circular ? Respuesta: 7! Extienda este resultado para n personas.

4. Debido a un error, 50 tornillos defectuosos fueron mezclados con 200 tornillos en buen estado. Si se venden 20 tornillos, ¿ Cuál es la probabilidad de que k sean defectuosos ? $(0 \leq k \leq 20)$.

5. Se lanza una moneda n veces. Supongamos el caso general $P(\text{"águila"}) = p$ y $P(\text{"sol"}) = q$, en donde $p, q > 0$ y $p+q = 1$. ¿ Cuál es la probabilidad de obtener k águilas ? $(0 \leq k \leq n)$.

6. Se lanzan simultaneamente cinco dados distinguibles. Demuestre que

   (a) $P(\text{"Todos los dados muestren caras diferentes"}) = 5/54$.

   (b) $P(\text{"Exactamente un par"}) = 25/54$.

   (c) $P(\text{"Exactamente dos pares diferentes"}) = 0.2315$.

   (d) $P(\text{"Exactamente tres caras iguales"}) = 25/162$.

   (e) $P(\text{"Exactamente tres caras iguales y un par"}) = 25/648$.

   (f) $P(\text{"Exactamente cuatro caras iguales"}) = 25/1296$.

   (g) $P(\text{"Cinco caras iguales"}) = 1/6^4$.

7. Se lanza una moneda n veces en donde p denota la probabilidad de obtener el resultado "sol" en cualquier lanzamiento. Demuestre que la probabilidad de obtener k soles por primera vez en el n-ésimo lanzamiento es

$$\binom{n-1}{k-1} p^k (1-p)^{n-k}.$$

8. Se lanza una moneda honesta tantas veces hasta que haya aparecido igual número de águilas que de soles.

   (a) ¿ Cuál es la probabilidad de que sea necesario lanzar la moneda dos veces ?

   (b) ¿ Cuál es la probabilidad de que sea necesario lanzar la moneda cuatro veces ?

(c) ¿ Cuál es la probabilidad de que sea necesario lanzar la moneda seis veces ?

(d) ¿ Cuál es la probabilidad de que sea necesario lanzar la moneda mas de seis veces ?

9. *Los cumpleaños.* Calcule la probabilidad de que en un conjunto de n personas, al menos dos de ellas tengan la misma fecha de cumpleaños. Sugerencia: Considere el complemento del evento de interés.

10. *La mesa circular.* Use el método de inducción para demostrar que las diferentes formas en que n personas pueden sentarse en una mesa circular es $(n-1)!$ Dibuje explicitamente los diferentes arreglos para $n = 3$ y luego $n = 4$.

11. *Las dos parejas.* Dos parejas se sientan al azar en una mesa lineal, ¿ Cuál es la probabilidad de que ninguna persona quede sentada junto a su pareja ?

12. *Las tres parejas.* Tres parejas se sientan al azar en una mesa lineal, ¿ Cuál es la probabilidad de que ninguna persona quede sentada junto a su pareja ?

13. *Las cuatro parejas.* Cuatro parejas se sientan al azar en una mesa lineal, ¿ Cuál es la probabilidad de que ninguna persona quede sentada junto a su pareja ?

14. *Paquetes de calcetines.* Para hacer rendir su bajo salario, cierto profesor de probabilidad compra paquetes de calcetines de cinco pares, todos iguales, para ser usados uno cada dia de la semana. Llegado el fin de semana el profesor mete a la lavadora los cinco pares, y una vez secos los acomoda nuevamente en pares para ser usados la semana siguiente.

    (a) ¿ Cuál es la probabilidad de que los calcetines hayan sido apareados el primer fin de semana como originalmente se compraron ?

(b) ¿ Cuál es la probabilidad de que los calcetines hayan sido apareados tanto el primero como el segundo fin de semana como originalmente se compraron ?

15. *Melate.* Cierto juego de la lotería nacional consiste en adivinar seis números que se escogen totalmente al azar del conjunto $\{1, 2, 3, \ldots, 39\}$. La selección se efectúa sin reemplazo y el orden no importa, es decir, se trata de combinaciones. Además de los seis números se escoge un número adicional para determinar los diferentes premios. ¿ Calcule la probabilidad de obtener

    (a) el primer lugar (6 naturales)?

    (b) el segundo lugar (6 naturales + adicional)?

    (c) el tercer lugar (5 naturales)?

    (d) el cuarto lugar (4 naturales)?

    (e) el quinto lugar (3 naturales + adicional)?

2.4 Probabilidad Geométrica

La probabilidad geométrica es una extensión de la definición clásica de la probabilidad: "Número de casos favorables entre número de casos totales". En la probabilidad geométrica el conteo de casos se sustituye por el cálculo de áreas y volumenes.

(Ver ejemplos de clase)

EJERCICIOS

1. Se escogen dos números x, y al azar dentro del intervalo unitario $[0, 1]$. ¿ Cuál es la probabilidad de que la suma de x y y sea mayor que uno y que al mismo tiempo la suma de sus cuadrados sea menor que uno ? Respuesta: $(\pi - 2)/4$.

2. Se escogen dos puntos a y b al azar dentro del intervalo unitario $[0, 1]$. ¿ Cuál es la probabilidad de que

   (a) la distancia de a a b sea mayor a $1/2$?

   (b) la distancia de a a b sea menor a $1/2$?

   (c) la distancia de a a 0 sea igual a la distancia de b a 1 ?

3. Una varilla de metal de longitud l se rompe en dos puntos distintos escogidos al azar. Demuestre que la probabilidad de que los tres segmentos asi obtenidos formen un triángulo es $1/4$.

4. Un tren y un autobús llegan a cierta estación de manera independiente y al azar entre las 9 a.m. y las 10 a.m. El tren se detiene en la estación durante 10 minutos y el autobús solamente x minutos. Encuentre el valor de x en minutos tal que la probabilidad de que el tren y el autobús se encuentren en la estación sea $1/2$. Respuesta: $x = 60 - \sqrt{1100}$.

5. *El problema de la aguja de Buffón.* Considere un conjunto infinito de lineas horizontales paralelas sobre una superficie plana. La distancia entre una linea y otra es d. Se deja caer una aguja de longitud l sobre la superficie. ¿ Cuál es la probabilidad de que la aguja no toque ninguna linea ?

6. Se escogen al azar y de manera independiente tres números a, b y c dentro del intervalo unitario $[0, 1]$. Demuestre que la probabilidad de que se cumpla la condición $a + b + c \leq 1$ es $1/6$.

2.5 Axiomatización de la Probabilidad

En esta sección expondremos la teoría axiomática de la probabilidad inventada por varios cientificos durante el primer tercio del siglo XX. No se trata de decir cómo deben calcularse las probabilidades, sino dar las reglas de cálculo que éstas deben satisfacer.

Primeramente definiremos el concepto de espacio de probabilidad, que es la base matemática de un experimento aleatorio.

Definición. *Un espacio de probabilidad es una terna* $(\Omega, \mathcal{F}, P)$ *en donde*

1. Ω *es un conjunto arbitrario.*
2. $\mathcal{F}$ *es una álgebra de subconjuntos de* Ω *y*
3. $P : \mathcal{F} \to R$ *es una medida de probabilidad.*

Explicaremos a continuación los tres elementos que constituyen un espacio de probabilidad. El primero de ellos es un conjunto Ω que corresponde al espacio muestral del experimento aleatorio de interés. El segundo elemento es un conjunto que denotamos por la letra $\mathcal{F}$ y que contiene a todos los eventos a los cuales deseamos calcular su probabilidad. Su definición formal es la siguiente.

Definición. *Sea* Ω *el espacio muestral de un experimento aleatorio. Decimos que una clase* $\mathcal{F}$ *de subconjuntos de* Ω *es una álgebra si se cumplen las siguientes tres condiciones*

1. $\Omega \in \mathcal{F}$.
2. *Si* $A \in \mathcal{F}$ *entonces* $A^c \in \mathcal{F}$.
3. *Si* A, $B \in \mathcal{F}$ *entonces* $A \cup B \in \mathcal{F}$.

Asi pues, la clase $\mathcal{F}$ contiene todos los eventos de interés. Para cada espacio muestral Ω, podemos asociar varias álgebras. He aqui algunos ejemplos de álgebras.

1. $\mathcal{F}_1 = \{\emptyset, \Omega\}$.
2. $\mathcal{F}_2 = \{\emptyset, A, A^c, \Omega\}$, con $A \subset \Omega$.
3. $\mathcal{F}_3 = 2^\Omega$.

El lector puede verificar directamente que tanto $\mathcal{F}_1$ como $\mathcal{F}_2$ y $\mathcal{F}_3$ satisfacen la definición de álgebra de subconjuntos de Ω. Observamos que $\mathcal{F}_1$ es el álgebra más pequeña y $\mathcal{F}_3$ es el álgebra más grande de subconjuntos de Ω. Estas cuestiones son objeto de estudio de la teoría de la probabilidad y de la teoría de la medida en general. Nosotros no pondremos mayor énfasis en estas cuestiones rigurosas pues este es un curso elemental de probabilidad.

Una vez que se establece el álgebra $\mathcal{F}$ de subconjuntos de Ω con la que se va a trabajar, a los elementos de $\mathcal{F}$ se les llama **eventos**. Para nuestro curso podremos suponer siempre que el álgebra $\mathcal{F}$ es el conjunto potencia 2^Ω. De modo que estamos considerando que todo subconjunto de Ω es un evento. Estamos ahora listos para establecer los axiomas de la probabilidad.

Axiomas de la Probabilidad (Kolmogorov, 1933)
Una medida de probabilidad P es una función

$$P : \mathcal{F} \to R$$

que satisface las siguientes tres condiciones

1. $P(\Omega) = 1$.
2. $P(A) \geq 0$ *para cualquier evento* A.
3. $P(A \cup B) = P(A) + P(B)$ *si* A *y* B *son ajenos.*

Con esto quedan entonces definidos todos los elementos que constituyen un espacio de probabilidad: un conjunto arbitrario Ω, un álgebra $\mathcal{F}$ de subconjuntos de Ω y una medida de probabilidad $P : \mathcal{F} \to [0,1]$.

Nos concentraremos ahora en el estudio de la función P. Basados en los anteriores axiomas y la teoría de conjuntos demostraremos ahora una serie de propiedades que satisface toda medida de probabilidad.

Proposición. $P(A^c) = 1 - P(A)$.

Demostración. Escribimos a Ω como la unión disjunta $\Omega = A \cup A^c$. Entonces

$$P(\Omega) = P(A) + P(A^c),$$

en donde por el primer axioma, $P(\Omega) = 1$. Despejando $P(A^c)$ obtenemos el resultado requerido. •

Proposición. $P(\emptyset) = 0$.

Demostración. Observamos que $\emptyset = \Omega^c$. Entonces

$$\begin{aligned} P(\emptyset) &= P(\Omega^c) \\ &= 1 - P(\Omega) \\ &= 0. \end{aligned}$$

•

Proposición. *Si* $A \subset B$ *entonces* $P(B - A) = P(B) - P(A)$.

Demostración. Como $A \subset B$, podemos escribir al conjunto B como la unión disjunta $B = A \cup (B - A)$. Entonces

$$P(B) = P(A) + P(B - A),$$

de donde se obtiene la propiedad requerida. •

Proposición. *Si* $A \subset B$ *entonces* $P(A) \leq P(B)$.

Demostración. Esta es una consecuencia de la propiedad anterior y del segundo axioma de la probabilidad.

$$P(B) = P(A) + P(B - A) \geq P(A).$$

•

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

Demostración. Observamos primero que $A \cup B = A \cup (B - A)$, en donde la unión del lado derecho es disjunta. Entonces por el tercer axioma de la probabilidad,

$$\begin{aligned} P(A \cup B) &= P(A) + P(B - A) \\ &= P(A) + P(B \cap A^c). \end{aligned} \tag{2.1}$$

Por otro lado $B = (B \cap A) \cup (B \cap A^c)$. Por lo tanto $P(B) = P(B \cap A) + P(B \cap A^c)$. De donde $P(B \cap A^c) = P(B) - P(B \cap A)$. Substituyendo en (2.1) tenemos el resultado requerido. •

Muchas otras propiedades de P pueden ser demostradas, algunas de ellas están contenidas en la siguiente sección de ejercicios.

EJERCICIOS

1. Sean A y B eventos tales que $P(A) = p$, $P(B) = q$ y $P(A \cap B) = r$. Demuestre que
   (a) $P(A \cap B^c) = p - r$.
   (b) $P(A^c \cap B^c) = 1 - (p + q - r)$.
   (c) $P(A \triangle B) = p + q - 2r$.
2. Demuestre la siguiente igualdad.

$$P(A \triangle B) = P(A) + P(B) - 2P(A \cap B).$$

3. Sea $A_1, A_2, \ldots, A_n$ una sucesión finita de eventos tales que $P(A_i) = 1$. Demuestre que

$$P\left(\bigcap_{i=1}^{n} A_i\right) = 1.$$

4. Sea $A_1, A_2, \ldots, A_n$ una sucesión finita de eventos tales que $P(A_i) = 0$. Demuestre entonces que

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 0.$$

5. ¿ Es posible que existan eventos A y B tales que $P(A) = 1/3$, $P(B) = 1/5$ y $P(A \cap B) = 1/4$?

6. Demuestre que para cualesquiera eventos A, B y C,

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad -P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad +P(A \cup B \cup C). \end{aligned}$$

Sugerencia: Escriba

$$P(A \cup B \cup C) = P[A \cup (B \cup C)],$$

y utilice la propiedad (e).

7. Proporcione un contraejemplo para las siguientes proposiciones.

   (a) Si $P(A) = P(B)$ entonces $A = B$.

   (b) Si $P(A \triangle B) = 0$ entonces $A = B$.

8. Demuestre las siguientes propiedades.

   (a) $P(A \cap B) \geq 1 - P(A^c) - P(B^c)$.

   (b) $P(A \cap B) \geq P(A) + P(B) - 1$.

(c) $P(A_1 \cap A_2 \cap \cdots \cap A_n) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$.

(d) $P(\emptyset) = 0$ sin usar $\Omega^c = \emptyset$.

(e) Si $A = B$ entonces $P(A \triangle B) = 0$.

(f) $P(A \cap B \cap C) \leq P(A \cap B) \leq P(A \cup B) \leq P(A \cup B \cup C)$.

(g) Si $P(A) > 1/2$ y $P(B) > 1/2$ entonces $P(A \cap B) > 0$.

(h) $P[(A \cap B^c) \cup (B \cap A^c)] = P(A) + P(B) - 2P(A \cap B)$.

9. Sean P y Q dos medidas de probabilidad definidas sobre la misma álgebra. Demuestre que si $\alpha \in [0,1]$ entonces $\alpha P + (1-\alpha)Q$ es también una medida de probabilidad.

10. Sean A y B eventos tales que $P(A) + P(B) = 1$. ¿ Bajo qué condiciones se tendrá que $A \cup B = \Omega$? Generalize su respuesta al caso de n eventos $A_1, A_2, \ldots, A_n$ tales que $P(A_1) + P(A_2) + \cdots + P(A_n) = 1$.

11. Sea $\Omega = R^2$ el espacio muestral de un experimento aleatorio. Sean los eventos

$$
\begin{aligned}
A_1 &= \{(x,y) \in \Omega : x \leq 5, y \leq 7\}. \\
A_2 &= \{(x,y) \in \Omega : x \leq 5, y \leq 1\}. \\
A_3 &= \{(x,y) \in \Omega : x \leq 3, y \leq 7\}. \\
A_4 &= \{(x,y) \in \Omega : x \leq 3, y \leq 1\}. \\
A_5 &= \{(x,y) \in \Omega : 3 < x \leq 5, 1 < y \leq 7\}.
\end{aligned}
$$

Supongamos que $P(A_1) = 3/4$, $P(A_2) = P(A_3) = 1/2$ y $P(A_4) = 3/8$. Dibuje Ω, los eventos A_1, A_2, A_3, A_4, A_5 y calcule $P(A_5)$.

12. Sean A, B y C tres eventos de un experimento aleatorio. Demuestre la siguiente igualdad.

$$
\begin{aligned}
P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\
&\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\
&\quad + P(A \cap B \cap C).
\end{aligned}
$$

Sugerencia: Escriba

$$P(A \cup B \cup C) = P\left[A \cup (B \cup C)\right],$$

y utilice la fórmula conocida para la probabilidad de la unión de dos eventos. Alternativamente el siguiente ejercicio nos provee de otro mecanismo para obtener la fórmula anterior.

13. Sean A, B y C tres eventos de un experimento aleatorio. Demuestre que la probabilidad de que exactamente uno de los eventos A, B o C ocurra es

$$\begin{aligned}&P(A) + P(B) + P(C)\\ &-2P(A \cap B) - 2P(A \cap C) - 2P(B \cap C)\\ &+3P(A \cap B \cap C).\end{aligned}$$

Ahora demuestre que la probabilidad de que exactamente dos de los eventos A, B o C ocurran es

$$P(A \cap B) + P(A \cap C) + P(B \cap C) - 3P(A \cap B \cap C).$$

Usando los dos cálculos anteriores demuestre que

$$\begin{aligned}P(A \cup B \cup C) \;=\;& P(A) + P(B) + P(C)\\ &-P(A \cap B) - P(A \cap C) - P(B \cap C)\\ &+P(A \cap B \cap C).\end{aligned}$$

2.6 Probabilidad Condicional

En esta sección estudiaremos el concepto de probabilidad condicional y demostraremos algunos resultados importantes relacionadose con este concepto. La probabilidad condicional nos va a permitir incorporar información adicional al cálculo de las probabilidades. Es un concepto muy sencillo y de amplio uso en las aplicaciones.

Definición. *Sea* $(\Omega, \mathcal{F}, P)$ *un espacio de probabilidad y* B *un evento tal que* $P(B) > 0$. *Definimos la* **probabilidad condicional** *de un evento* A *cualquiera dado el evento* B, *denotada por* $P(A|B)$, *como sigue*

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

El evento B representa información adicional que sabemos del experimento aleatorio. Sabiendo que el evento B ha ocurrido, las probabilidades de todos los eventos se han entonces modificado. La probabilidad modificada es la probabilidad condicional.

La probabilidad condicional se puede interpretar como sigue. Si ha ocurrido el evento B entonces no puede ocurrir algo fuera de B, de modo que el espacio muestral Ω se ha reducido al conjunto B. Para calcular la nueva probabilidad de un evento A, es necesario saber cuanto de A está también en B, i.e. $P(A \cap B)$ y dividirlo por la nueva totalidad, i.e. $P(B)$. El siguiente resultado es bastante útil.

Proposición. *Sea* $P(\cdot)$ *una medida de probabilidad y* B *un evento tal que* $P(B) > 0$. *Entonces* $P(\cdot|B)$ *es una medida de probabilidad, es decir, satisface los siguientes axiomas*

1. $P(\Omega|B) = 1$.
2. $P(A|B) \geq 0$ *para cualquier evento* A.
3. $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$ *si* A_1 *y* A_2 *son ajenos.*

El lector podrá comprobar fácilmente las anteriores tres igualdades, simplemente usando la definición de probabilidad condicional.

Nuestro resultado anterior tiene como consecuencia importante el hecho de que como $P(\cdot|B)$ es una medida de probabilidad, toda propiedad válida para $P(\cdot)$ es también válida para $P(\cdot|B)$.

Enunciaremos ahora algunos resultados importantes de la probabilidad condicional. Antes de ello necesitamos explicar lo que entenderemos por una partición de un conjunto.

Definición. *Sea Ω un conjunto arbitrario. Una* **partición** *de Ω es una colección de subconjunto $B_1, B_2, \ldots, B_n$ de Ω que satisface las tres condiciones siguientes*

1. $B_i \neq \emptyset$ *para* $i = 1, 2, \ldots, n$.
2. $B_i \cap B_j = \emptyset$ *si* $i \neq j$.
3. $\displaystyle\bigcup_{i=1}^{n} B_i = \Omega$.

Teorema de Probabilidad Total
Sea Ω un espacio muestral y $B_1, B_2, \ldots, B_n$ una partición de Ω tal que $P(B_i) > 0$. Entonces para cualquier evento A,

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Demostración. Tenemos que

$$\begin{aligned} A &= A \cap \Omega \\ &= A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) \\ &= \bigcup_{i=1}^{n} (A \cap B_i). \end{aligned}$$

Entonces por el tercer axioma de la probabilidad

$$P(A) \;=\; P\left(\bigcup_{i=1}^{n} (A \cap B_i)\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} P(A \cap B_i) \\
&= \sum_{i=1}^{n} P(A|B_i)P(B_i).
\end{aligned}
$$

•

En particular, cuando la partición consiste de los eventos B y B^c entonces el teorema de probabilidad total toma la expresión

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

Ahora ilustraremos la utilización de este resultado.

Ejemplo. (La urna de Polya) Suponga que en una urna se tienen r bolas rojas y b bolas blancas. El experimento consiste en seleccionar una bola al azar y regresarla a la urna junto con c bolas del mismo color. Sea R_n el evento "Se selecciona una bola roja en la n-ésima extracción". Calcule $P(R_1)$ y $P(R_2)$.

Solución. Claramente $P(R_1) = \dfrac{r}{r+b}$. Para calcular $P(R_2)$ usaremos el teorema de probabilidad total.

$$
\begin{aligned}
P(R_2) &= P(R_2|R_1)P(R_1) + P(R_2|R_1^c)P(R_1^c) \\
&= \frac{r+c}{r+b+c}\,\frac{r}{r+b} + \frac{r}{r+b+c}\,\frac{b}{r+b} \\
&= \frac{r(r+b+c)}{(r+b+c)(r+b)} \\
&= \frac{r}{r+b}.
\end{aligned}
$$

Observamos que se obtiene el mismo resultado que antes. Puede demostrarse por inducción que

$$P(R_n) = r/(r+b).$$

Teorema de Bayes
Sea Ω un espacio muestral y $B_1, B_2, \ldots, B_n$ una partición de Ω tal que $P(B_i) > 0$. Entonces para cualquier evento A con $P(A) > 0$ y cualquier $j = 1, 2, \ldots, n$ tenemos que

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\displaystyle\sum_{i=1}^{n} P(A|B_i)P(B_i)}.$$

Demostración. Este resultado es una consecuencia de la definición de probabilidad condicional y del teorema de probabilidad total. Tenemos que

$$\begin{aligned} P(B_j|A) &= \frac{P(B_j \cap A)}{P(A)} \\ &= \frac{P(A|B_j)P(B_j)}{P(A)} \\ &= \frac{P(A|B_j)P(B_j)}{\displaystyle\sum_{i=1}^{n} P(A|B_i)P(B_i)}. \end{aligned}$$

•

Ejemplo. Suponga que en una cierta población el 70% son hombres y 30% son mujeres. Suponga también que el 50% de las mujeres fuman y el 40% de los hombres fuman. Se escoge a una persona al azar. ¿ Cuál es la probabilidad de que esta persona que sabemos que fuma sea hombre ?

Solución. Sean los eventos

$$\begin{aligned} F &= \text{"La persona fuma"}, \\ H &= \text{"La persona es hombre"}. \end{aligned}$$

La pregunta planteada se traduce entonces al cálculo de la probabilidad $P(H|F)$. Calcularemos $P(H|F)$ usando el teorema de Bayes.

$$\begin{aligned} P(H|F) &= \frac{P(F|H)P(H)}{P(F|H)P(H)+P(F|H^c)P(H^c)} \\ &= \frac{(0.4)(0.7)}{(0.4)(0.7)+(0.5)(0.3)} = 0.65 \end{aligned}$$

Regla de Multiplicación
Sean $A_1, A_2, \ldots, A_n$ eventos tales que $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$. Entonces

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_n) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots \\ &\quad \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \end{aligned}$$

Demostración. Esta proposición es evidente. Para demostrarla únicamente necesitamos substituir la definición de probabilidad condicional en el lado derecho y cancelar varios términos para obtener el lado izquierdo. •

Ejemplo. ¿ Cuál es la probabilidad de que tres bolas rojas sean obtenidas en las tres primeras extracciones de la urna de Polya ?

Solución. Recordemos que en el problema de la urna de Polya, R_n denota el evento "Se selecciona una bola roja en la n-ésima extracción". Nos preguntamos entonces por la probabilidad del evento $R_1 \cap R_2 \cap R_3$. Por la regla de multiplicación tenemos que

$$\begin{aligned} P(R_1 \cap R_2 \cap R_3) &= P(R_1)P(R_2|R_1)P(R_3|R_1 \cap R_2) \\ &= \left(\frac{r}{r+b}\right)\left(\frac{r+c}{r+b+c}\right)\left(\frac{r+2c}{r+b+2c}\right). \end{aligned}$$

EJERCICIOS

1. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y B un evento tal que $P(B) > 0$. Demuestre que $P(\cdot|B)$ es una medida de prob-

abilidad. Se sigue entonces que $(\Omega, \mathcal{F}, P(\cdot|B))$ es un espacio de probabilidad.

2. Decimos que un evento A favorece a otro evento B si $P(B|A) > P(B)$. Si A favorece a B y B favorece a C, ¿ Se sigue que A favorece a C ?

3. Diga falso o verdadero. En cada caso demuestre o proporcione un contraejemplo.

   (a) Si $P(A|B) > P(A)$ entonces $P(B|A) > P(B)$.

   (b) Si $P(A) \geq P(B)$ entonces $P(A|C) \geq P(B|C)$.

4. Suponga $P(B) = P(A|B) = P(C|A \cap B) = p$. Demuestre que $P(A \cap B \cap C) = p^3$.

5. Una persona toma al azar uno de los números 1, 2 o 3, y luego tira un dado tantas veces como indica el número. Sumando el resultado de las tiradas, ¿ Cuál es la probabilidad de que obtenga un total de 5 ?

6. *La paradoja de la caja de Bertrand.* Se tienen 3 cajas y cada una de ellas tiene 2 monedas. La caja C_1 contiene dos monedas de oro. La caja C_2 contiene una moneda de oro y una de plata. La caja C_3 contiene dos monedas de plata. Se selecciona una caja al azar y de alli se escoge una moneda. Si la moneda es de oro, ¿ Cuál es la probabilidad de que provenga de la caja con dos monedas de oro ? La respuesta no es 1/2 sino 2/3.

7. Considere el modelo de la urna de Polya. Demuestre por inducción sobre n que $P(R_n) = r/(r+b)$.

8. *El problema de los tres prisioneros.* A tres prisioneros, a quienes podemos llamar A, B y C, les informa su celador que se ha escogido al azar a uno de ellos para ser ejecutado y que los otros dos quedarán en libertad. El prisionero A razona que tiene 1/3 de probabilidad de ser ejecutado. A continuación le pide al celador que le diga en secreto cuál de

sus dos compañeros saldrá en libertad, argumentando que no tiene nada de malo divulgar esta información puesto que él ya sabe de antemano que por lo menos uno de ellos será liberado y que por otra parte, el saberlo no afectará su probabilidad de ser ejecutado. El celador, que es una persona ética, le señala por el contrario que, si A supiera cuál de sus compañeros será puesto en libertad, entonces su probabilidad de ser ejecutado aumentaría a $1/2$, pues el prisionero A sería uno de los dos prisioneros de los cuales uno va a ser ejecutado.

Suponiendo que en el caso de que A vaya a ser ejecutado, el celador contestara al azar que B o C será puesto en libertad. ¿ Quién tiene la razón, el prisionero o el celador ? Justifique su respuesta.

2.7 Independencia de Eventos

El concepto de independencia distingue a la probabilidad de otras ramas de las matemáticas. Es un concepto muy importante y surge directamente del vínculo de la probabilidad y la vida real.

Definición. *Decimos que dos eventos A y B son* **independientes** *si y solo si cumplen la condición*

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

De manera equivalente, la anterior condición puede escribirse como $P(A|B) = P(A)$ cuando $P(B) \neq 0$. En este caso la interpretación es sencilla: la ocurrencia del evento B no afecta la probabilidad del evento A, son independientes. Lo mismo puede decirse de la condición equivalente $P(B|A) = P(B)$. Más generalmente tenemos la siguiente definición.

Definición. *Decimos que n eventos $A_1, A_2, \ldots, A_n$ son* **independientes** *si cumplen con todas y cada una de las siguientes condi-*

ciones

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j) \quad i,j \textit{ distintos,} \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k) \quad i,j,k \textit{ distintos,} \\ &\vdots \\ P(A_1 \cap \cdots \cap A_n) &= P(A_1)\cdots P(A_n). \end{aligned}$$

Esto significa que para verificar la independencia de n eventos es necesario verificar la independencia dos a dos, tres a tres, etc., pues resulta que la independencia por pares no implica la independencia tres a tres, ni tampoco independencia tres a tres implica independencia por pares. En la sección de ejercicios se pide demostrar esta afirmación. De esta forma, para demostrar la independencia de n eventos realmente se necesitan demostrar todas las igualdades de nuestra definición, que son un total de $2^n - n - 1$ igualdades.

EJERCICIOS

1. Demuestre que para cualquier evento A,
   (a) A y $\emptyset$ son independientes.
   (b) A y Ω son independientes.
2. Demuestre que las siguientes cuatro afirmaciones son equivalentes.
   (a) A y B son independientes.
   (b) A y B^c son independientes.
   (c) A^c y B son independientes.
   (d) A^c y B^c son independientes.
3. Sean A yB dos eventos independientes tales que $P(A) = p_1$ y $P(B) = p_2$. Calcule la probabilidad de que no ocurre ninguno de estos eventos.

4. Las siguientes dos afirmaciones son en general falsas. Proporcione un contraejemplo para cada una de ellas.

   (a) "Si A y B son independientes entonces A y B son ajenos".

   (b) "Si A y B son ajenos entonces A y B son independientes".

5. Sea A un evento que es independiente consigo mismo. Demuestre entonces que $P(A) = 0$ o $P(A) = 1$.

6. Sea A un evento tal que $P(A) = 0$. Demuestre que A es independiente con cualquier otro evento B.

7. Sea A un evento tal que $P(A) = 1$. Demuestre que A es independiente con cualquier otro evento B.

8. Sean A, B y C eventos mutuamente independientes. Demuestre que $A \cup B$ y C son independientes. Proporcione un contraejemplo para el recíproco.

9. ¿ Cuántas igualdades son necesarias verificar para demostrar que n eventos son independientes ? Respuesta : $2^n - 1 - n$. Utilice el teorema del binomio.

10. Sean $A_1, A_2, \ldots, A_n$ eventos independientes. Demuestre que

$$P(\bigcup_{i=1}^{n} A_i) = 1 - \prod_{i=1}^{n} P(A_i^c).$$

11. Sean A, B y C tres eventos. Demuestre que

    (a) independencia dos a dos no implica independencia tres a tres.

    (b) independencia tres a tres no implica independencia dos a dos.

12. Sean $A_1, A_2, \ldots, A_n$ eventos mutuamente independientes tales que $P(A_i) = p$, $i = 1, 2, \ldots, n$. Calcule la probabilidad de que

(a) al menos uno de los eventos ocurra.

(b) exactamente m de los eventos ocurran.

(c) al menos m de los eventos ocurran.

13. Sea $\Omega = [0,1]$. Si a conjuntos de la misma longitud se les asigna la misma probabilidad, encuentre condiciones suficientes y necesarias para que dos eventos sean independientes.

14. Sean A y B eventos independientes tales que $P(A) > 0$ y $P(B) > 0$. Demuestre que entonces A y B no son ajenos. Ilustre este resultado con un ejemplo.

15. Demuestre o proporcione un contraejemplo para las siguientes proposiciones.

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

    (b) Si $P(B|A^c) = P(B|A)$ entonces A y B son independientes.

16. Supongamos que cierto experimento aleatorio tiene únicamente dos posibles resultados que llamaremos A y B. Realizamos una sucesión de ensayos independientes de este experimento hasta que se obtienen dos resultados consecutivos iguales.

    (a) Determine un posible espacio muestral para este experimento.

    (b) Suponga que $P(A) = p$ y $P(B) = 1 - p$. ¿ Cuál es la probabilidad de que el experimento termine en el n-ésimo ensayo ? $(n \geq 2)$.

17. Considere un experimento aleatorio con espacio muestral $\Omega = [0,1]$. Definimos la probabilidad de un intervalo $[a,b] \subset [0,1]$ como $P[a,b] = b - a$. Encuentre eventos A y B tales que

    (a) sean independientes y ajenos.

    (b) sean independientes pero no ajenos.

(c) sean ajenos pero no independientes.

(d) sean no ajenos y no independientes.

Capítulo 3

Variables Aleatorias

3.1 Variables Aleatorias

El concepto de variable aleatoria es un artificio que nos va a permitir traducir los posibles resultados de un experimento aleatorio en números reales.

Definición. *Dado un espacio de probabilidad* $(\Omega, \mathcal{F}, P)$, *una* **variable aleatoria** *es una función* $X : \Omega \to R$ *tal que para todo* $x \in R$, *se cumple la condición*

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}.$$

En el presente curso no nos ocuparemos de la condición técnica arriba mencionada. En cursos más avanzados de probabilidad y en teoría de la medida, a esta condición se le conoce como la medibilidad de la función X . De manera informal nosotros consideraremos simplemente que una variable aleatoria es una función $X : \Omega \to R$. Usualmente se utilizan las últimas letras del alfabeto en mayúsculas

X, Y, Z, para denotar a las variables aleatorias, es decir, a las funciones. Y para denotar un valor general tomado por estas variables aleatorias usaremos las mismas letras pero en minúsculas x, y, z.

Damos a continuación algunos ejemplos de variables aleatorias.

Ejemplo. Consideremos el experimento de lanzar una moneda. Entonces tenemos que $\Omega = \{\text{águila}, \text{sol}\}$. Podemos entonces definir la variable aleatoria X como sigue

$$\begin{aligned} X(\text{águila}) &= 0, \\ X(\text{sol}) &= 1. \end{aligned}$$

De esta forma podemos entonces pensar que los posibles resultados de nuestro experimento aleatorio son los números 0 y 1, escogidos por nosotros de manera arbitraria. El lector podrá escoger otro par de números y definir otra variable aleatoria. Observamos además que los únicos valores que puede tomar la v.a. X son el 0 y el 1, y ningún otro valor. En otras palabras, el rango de la función X es el conjunto $\{0, 1\}$. Si la moneda lanzada es honesta podemos además establecer que

$$\begin{aligned} P(X = 0) &= 1/2, \\ P(X = 1) &= 1/2. \end{aligned}$$

Ejemplo. Consideremos el experimento aleatorio de escoger al azar un número real del intervalo unitario $[0, 1]$. En este caso es obvio que $\Omega = [0, 1]$, es decir, los resultados de este experimento aleatorio son ya números reales. Podemos definir la variable aleatoria X simplemente como la función identidad

$$X(\omega) = \omega.$$

Definición. *Decimos que una variable aleatoria $X : \Omega \to R$ es* **discreta** *si el rango de la función X es un conjunto finito o numerable. Si no es asi, si el rango de X es un conjunto infinito* *no numerable entonces decimos que X es una variable aleatoria* **continua**.

Sólo consideraremos variables aleatorias de un tipo o de otro, pero no una combinación de ambos tipos. Nuestro primer ejemplo de v.a. es del tipo discreto, y el segundo ejemplo es de tipo continuo. En general, vamos a suponer que una función cualquiera de una o varias variables aleatorias sigue siendo una variable aleatoria. Por ejemplo, si X y Y son variables aleatorias entonces también lo serán X^2, $\sin(X)$, $X+Y$, $X-Y$, $X \cdot Y$, X/Y, etc., cuando estas operaciones esten bien definidas. El término "variable aleatoria" será de uso frecuente en el resto de nuestro curso.

EJERCICIOS

1. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y sea $A \in \mathcal{F}$ un evento cualquiera. Definimos la **función indicadora** del evento A, denotada por $I_A : \Omega \to \{0,1\}$, como sigue

$$I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

   Demuestre que I_A es una variable aleatoria.

2. Sea c una constante. Demuestre que la función constante $X = c$ es una variable aleatoria.

3. Un experimento consiste en escoger al azar un punto del disco unitario con centro en el origen del plano Cartesiano. Tenemos entonces que el espacio muestral es el conjunto

$$\Omega = \{(x,y) \in \mathbf{R^2} : x^2 + y^2 \leq 1\}.$$

   Sea X la variable aleatoria dada por la proyección $X(x,y) = x$. Calcule las siguientes probabilidades.

   (a) $P(X \in [-1/2, 1/2])$.

(b) $P(X \geq 1/3)$.

(c) $P(X^2 \geq 1/4)$.

4. Sea $I_A : \Omega \to \{0, 1\}$ la función indicadora del evento A. Demuestre las siguientes propiedades.

   (a) $I_\Omega = 1$.

   (b) $I_\emptyset = 0$.

   (c) $I_{\cup_i A_i} = \sum_i I_{A_i}$.

   (d) $I_{\cap_i A_i} = \prod_i I_{A_i}$.

   (e) $I_{A \cup B} = I_A + I_B - I_{A \cap B}$.

3.2 Funciones de Densidad y de Distribución

Definición. *Sea X una variable aleatoria discreta que toma los valores $x_0, x_1, x_2, \ldots, x_n, \ldots$ Definimos la* **función de densidad** *de X, y la denotamos por $f_X(x) : \mathbf{R} \to \mathbf{R}$, como sigue*

$$f_X(x) = \begin{cases} P(X = x) & \text{si } x = x_0, x_1, x_2, \ldots, x_n, \\ 0 & \text{otro caso.} \end{cases}$$

Observamos que la función $f_X(x)$ tiene como subindice el nombre de la v.a. X, y el argumento es la letra minúscula x. Observamos también que la letra f es minúscula. Estas observaciones aparentemente sin trascendencia son importantes de recordar. Para variables aleatorias continuas tenemos la siguiente definición.

Definición. *Sea X una variable aleatoria continua. Definimos la* **función de densidad** *de X como aquella función continua $f_X : \mathbf{R} \to \mathbf{R}$ tal que para todo intervalo $[a, b]$ de $\mathbf{R}$,*

$$P(X \in [a, b]) = \int_a^b f_X(x)\, dx.$$

Es decir, el área bajo la función de densidad en un intervalo $[a, b]$ es la probabilidad de que la variable aleatoria tome un valor dentro del intervalo $[a, b]$. Tanto en el caso discreto como en el caso continuo, toda función de densidad satisface las siguiente dos propiedades.

Proposición. *Sea $f_X(x)$ la función de densidad de la variable aleatoria continua X. Entonces se cumplen las siguientes propiedades.*

1. $f_X(x) \geq 0$.
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

Para el caso de variables aleatorias discretas, se tienen propiedades análogas. La única diferencia radica en substituir la integral por la suma sobre todos los posible valores que puede tomar la variable aleatoria X. De hecho, las propiedades arriba mencionadas definen lo que entederemos por función de densidad de una manera general.

Definición. *Decimos que una función continua $f : R \to R$ es de* **densidad** *si cumple las dos condiciones siguientes.*

1. $f(x) \geq 0$.
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1$.

Tanbién llamaremos función de densidad a toda función $f : R \to R$ que toma un número discreto de valores y que satisface las dos condiciones arriba mencionadas, la segunda de ellas adecuadamente escrita en términos de una suma.

Además de la función de densidad f_X que podemos asociarle a casi toda variable aleatoria X, existe otra función más importante aún desde el punto de vista matemático. Esta es la función de distribución que estudiaremos a continuación.

Definición. *Dada una variable aleatoria X, ya sea discreta o continua, definimos la* **función de distribución** *de X, denotada por $F_X(x) : \mathbf{R} \to \mathbf{R}$, como*

$$f_X(x) = P(X \leq x).$$

Nuevamente observamos que el nombre de la v.a. X aparece como subindice de la letra mayúscula F y que el argumento de la función es la letra minúscula x. La relación existente entre la función de densidad f_X y la función de distribución F_X de una variable aleatoria continua X se expresa entonces como sigue

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)\, du,$$

con una expresión análoga en el caso discreto. A la función de distribución F_X se le conoce también con el nombre de **función de acumulación de probabilidad**, resultando este término natural pues $F_X(x)$ es la probabilidad acumulada hasta el valor x. He aqui algunas propiedades de toda función de distribución.

Proposición. *Sea X una variable aleatoria cualquiera con función de distribución $F_X(x)$. Entonces*

1. $\lim_{x \to \infty} F_X(x) = 1.$
2. $\lim_{x \to -\infty} F_X(x) = 0.$
3. *Si $a \leq b$ entonces $F_X(a) \leq F_X(b)$.*
4. *$F_X(x)$ es una función continua por la derecha.*

Definición. *Decimos que una función cualquiera $F : R \to R$ es una función de distribución, si cumple con las cuatro propiedades arriba mencionadas.*

¿ Cómo encontramos $F_X(x)$ a partir de $f_X(x)$?

Conociendo la función de densidad $f_X(x)$, podemos encontrar al función de distribución $F_X(x)$ simplemente integrando en el caso continuo o sumando en el caso discreto. Para el caso continuo tenemos

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du.$$

En el caso discreto se suman todos los valores de $f_X(u)$ para valores de u menores o iguales a x.

¿ Cómo encontramos $f_X(x)$ a partir de $F_X(x)$?

En el caso de variables aleatorias continuas sabemos que $f_X(x)$ y $F_X(x)$ guardan la relación

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du.$$

Por el teorema fundamental del cálculo tenemos entonces que

$$f_X(x) = \frac{d}{dx} F_X(x).$$

Para el caso de variables discretas usamos la relación

$$f_X(x) = F_X(x+) - F_X(x-).$$

Es decir, $f_X(x)$ es el tamaño de la discontinuidad de F_X en el punto x.

EJERCICIOS

1. Sean $f_1(x)$ y $f_2(x)$ dos funciones de densidad y $\alpha \in [0,1]$ una constante. Demuestre que $\alpha f_1(x) + (1-\alpha) f_2(x)$ es una función de densidad.

2. Considere el experimento aleatorio de lanzar una moneda honesta dos veces consecutivas. Determine el espacio muestral para este experimento. Defina ahora la variable aleatoria X como el número de "águilas" obtenidas en los dos lanzamientos.

   (a) Calcule y grafique f_X.

   (b) Calcule y grafique F_X.

   (c) Calcule $P(X \geq 1)$.

3. La variable aleatoria X tiene la siguiente función de densidad

$$f(x) = \begin{cases} \frac{3}{2}x^2 & \text{si } -1 < x < 1. \\ 0 & \text{otro caso.} \end{cases}$$

   Grafique $f(x)$ y calcule las siguientes probabilidades. Muestre estas probabilidades gráficamente sombreando las correspondientes áreas bajo la función de densidad.

   (a) $P(-1/4 < X < 2/3)$.

   (b) $P(|X| < 1/2)$.

   (c) $P(X \in (-3/4, -1/4) \cap (-1/2, 1/2))$.

4. Determine el valor de la constante c para cada una de las siguientes funciones de densidad discretas.

   (a) $f(x) = \begin{cases} cx^2 & \text{si } x = 0, 1, 2, \ldots, n. \\ 0 & \text{otro caso.} \end{cases}$

   (b) $f(x) = \begin{cases} c(1/4)^x & \text{si } x = 0, 1, 2, \ldots \\ 0 & \text{otro caso.} \end{cases}$

   (c) $f(x) = \begin{cases} ce^{-|x|} & \text{si } x = 0, \pm 1, \pm 2, \ldots \\ 0 & \text{otro caso.} \end{cases}$

5. Determine el valor de la constante c para cada una de las siguientes funciones de densidad continuas. Grafique en cada caso.

(a) $f(x) = \begin{cases} c|x| & \text{si } x \in [-1, 1]. \\ 0 & \text{otro caso.} \end{cases}$

(b) $f(x) = \begin{cases} c(1 + \sin x) & \text{si } x \in [0, 2\pi]. \\ 0 & \text{otro caso.} \end{cases}$

(c) $f(x) = ce^{-|x|}$.

6. Sea X una variable aleatoria discreta con función de distribución dada por

$$F_(x) = \begin{cases} 0 & \text{si } x < -6. \\ 1/4 & \text{si } -6 \leq x < -3. \\ 1/2 & \text{si } -3 \leq x < 3. \\ 3/4 & \text{si } 3 \leq x < 6. \\ 1 & \text{si } x \geq 6. \end{cases}$$

(a) Grafique $F_X(x)$.

(b) Obtenga y grafique $f_X(x)$.

(c) Calcule las siguientes probabilidades: $P(X \leq 3)$, $P(X = 3)$, $P(X < 3)$, $P(X \geq 1)$, $P(-4 < X < 4)$, $P(X = 5)$.

7. Grafique y demuestre que las siguientes funciones son funciones de densidad. Encuentre y grafique las correspondientes funciones de distribución.

(a) $f(x) = \begin{cases} 0.3 & \text{si } x = -5. \\ 0.2 & \text{si } x = 2, -3. \\ 0.3 & \text{si } x = 4. \\ 0 & \text{otro caso.} \end{cases}$

(b) $f(x) = \begin{cases} x/15 & \text{si } x = 1, 2, 3, 4, 5. \\ 0 & \text{otro caso.} \end{cases}$

(c) $f(x) = \begin{cases} \frac{3}{2}\sqrt{x} & \text{si } 0 < x < 1. \\ 0 & \text{otro caso.} \end{cases}$

(d) $f(x) = \begin{cases} |x| & \text{si } -1 \leq x \leq 1. \\ 0 & \text{otro caso.} \end{cases}$

(e) $f(x) = \begin{cases} \frac{3}{2}(1+x)^2 & \text{si } -1 < x < 0. \\ \frac{3}{2}(1-x)^2 & \text{si } 0 < x < 1. \\ 0 & \text{otro caso.} \end{cases}$

8. Grafique y demuestre que las siguientes funciones son funciones de distribución. Encuentre y grafique las correspondientes funciones de densidad.

(a) $F(x) = \begin{cases} 0 & \text{si } x < 1. \\ 1/3 & \text{si } 1 \leq x < 2. \\ 1 & \text{si } x \geq 2. \end{cases}$

(b) $F(x) = \begin{cases} 0 & \text{si } x < 0. \\ \sqrt{x} & \text{si } 0 \leq x < 1. \\ 1 & \text{si } x \geq 1. \end{cases}$

(c) $F(x) = \begin{cases} 0 & \text{si } x < 0. \\ x^2 & \text{si } 0 \leq x < 1/2. \\ 3x/2 - 1/2 & \text{si } 1/2 \leq x < 1. \\ 1 & \text{si } x \geq 1. \end{cases}$

(d) $F(x) = \begin{cases} 0 & \text{si } x < -2. \\ (x+2)/2 & \text{si } -2 \leq x < -1. \\ 1/2 & \text{si } -1 \leq x < 1. \\ x/2 & \text{si } 1 \leq x < 2. \\ 1 & \text{si } x \geq 2. \end{cases}$

9. Sea X una variable aleatoria continua con función de densidad dada por

$$f_X(x) = \begin{cases} -kx^3 & \text{si } -1 \leq x < 0. \\ kx^3 & \text{si } 0 \leq x < 1. \\ 0 & \text{otro caso.} \end{cases}$$

(a) Obtenga el valor de k y grafique $f_X(x)$.

(b) Obtenga y grafique $F_X(x)$.

(c) Obtenga el valor α tal que $P(-\alpha \leq X \leq \alpha) = 1/2$ y muestre gráficamente esta probabilidad.

3.3 Distribuciones Discretas

A continuación estudiaremos algunas distribuciones de probabilidad de variables aleatorias discretas de uso común.

Distribución Uniforme Discreta

Decimos que una variable aleatoria X tiene una distribución uniforme discreta sobre el conjunto de números $\{x_1, x_2, \ldots, x_n\}$ si la probabilidad de que X tome cualquiera de estos valores es la misma, es decir, $1/n$. Esta distribución surge en espacios de probabilidad equiprobables, esto es, en situaciones en donde tenemos n resultados diferentes y todos ellos tienen la misma probabilidad de ocurrencia. Los juegos de lotería son un ejemplo donde puede aplicarse esta distribución de probabilidad.

Escribimos entonces $X \sim \text{Unif}\{x_1, x_2, \ldots, x_n\}$ si

$$P(X = x) = \begin{cases} 1/n & \text{si } x = x_1, x_2, \ldots, x_n. \\ 0 & \text{otro caso.} \end{cases}$$

Distribución Bernoulli

Un ensayo Bernoulli se define como aquel experimento aleatorio con únicamente dos posibles resultados, éxito o fracaso, con probabilidades respectivas p y $1 - p$. Si definimos la variable aleatoria X como

$$X(\text{éxito}) = 1,$$
$$X(\text{fracaso}) = 0,$$

entonces decimos que X tiene una distribución Bernoulli con parámetro $p \in (0,1)$. Y escribimos $X \sim \mathrm{Ber}(p)$ si

$$P(X = x) = \begin{cases} p^x(1-p)^{1-x} & \text{si } x = 0, 1. \\ 0 & \text{otro caso.} \end{cases}$$

En la realización de todo experimento aleatorio siempre es posible preguntarnos por la ocurrencia, o no, de un evento cualquiera. Por ejemplo ganar o no ganar en un juego de lotería. Este es el esquema general donde surge esta distribución, que aunque sencilla, es de amplia aplicación.

Distribución Binomial

Supongamos ahora que tenemos una serie de n ensayos independientes Bernoulli en donde la probabilidad de éxito en cualesquiera de estos n ensayos es siempre la misma probabilidad p. En este caso el experimento aleatorio consiste en realizar sucesivamente los n ensayos Bernoulli. Si denotamos por E el resultado "éxito" y por F el resultado "fracaso" entonces el espacio muestral consiste de todas las posibles sucesiones de tamaño n de caracteres E y F. Esto es,

$$\begin{aligned} \Omega = \{ \quad & EE\cdots EE, \\ & FE\cdots EE, \\ & \vdots \\ & FF\cdots FF\}. \end{aligned}$$

Usando el principio multiplicativo, es fácil ver el conjunto Ω tiene 2^n elementos. Si ahora definimos la variable aleatoria X como aquella que cuenta el número de éxitos en cada una de estas sucesiones, esto es

$$X(EE\cdots EE) \quad = \quad n,$$

$$
\begin{aligned}
X(FE\cdots EE) &= n-1,\\
&\vdots\\
X(FF\cdots FF) &= 0,
\end{aligned}
$$

entonces tenemos que X puede tomar los valores $0, 1, 2, \ldots, n$ con ciertas probabilidades que mencionaremos más adelante. Decimos entonces que X tiene una distribución binomial con parámetros n y p. Y escribimos $X \sim \text{bin}(n, p)$ si

$$
P(X = x) = \begin{cases} \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, \ldots, n. \\ 0 & \text{otro caso.} \end{cases}
$$

La fórmula anterior puede justificarse de la forma siguiente. Queremos que en n ensayos Bernoulli se obtengan x éxitos y $n - x$ fracasos. La probabilidad de obtener ésto es el número

$$
\underbrace{p \cdots p}_{x} \cdot \underbrace{(1-p) \cdots (1-p)}_{n-x} = p^x (1-p)^{n-x},
$$

pero hemos colocado los x éxitos en los primeros x ensayos, tenemos que multiplicar por las diferentes formas en que estos x éxitos pueden distribuirse en los n ensayos, este factor es el coeficiente binomial $\left(\begin{array}{c} n \\ x \end{array} \right)$.

Distribución Geométrica

Supongamos nuevamente que tenemos una sucesión de ensayos independientes Bernoulli, pero esta vez tenemos una sucesión infinita. Para cada una de los resultados de esta sucesión infinita definimos la variable aleatoria X como el número de fracasos antes de obtener el primer éxito. Por ejemplo,

$$\begin{aligned} X(FEFEFF\cdots) &= 1, \\ X(EFFEEE\cdots) &= 0, \\ X(FFFEFE\cdots) &= 3. \end{aligned}$$

Observamos entonces que X puede tomar los valores $0, 1, 2, \ldots$. La probabilidad de que X tome el valor x es $p(1-p)^x$. Decimos entonces que X tiene una distribución geométrica con parámetro p. Y escribimos $X \sim \text{Geo}(p)$ cuando

$$P(X = x) = \begin{cases} p(1-p)^x & \text{si } x = 0, 1, 2, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

El nombre de esta distribución proviene del hecho de que cuando escribimos la suma de todas las probabilidades, obtenemos una suma geométrica. La inspección sucesiva de artículos, posiblemente para control de calidad, puede modelarse usando una distribución geométrica.

Distribución Poisson

Supongamos que deseamos observar el número de ocurrencias de un cierto evento dentro de un intervalo de tiempo dado. Por ejemplo, el número de clientes que llegan a un cajero automático durante la noche, o tal vez deseamos registrar el número de accidentes que ocurren en cierta avenida durante todo un dia. Para modelar este tipo de situaciones definimos la variable aleatoria X como el número de ocurrencia de estos eventos en el intervalo de tiempo dado.

Es claro entonces que X puede tomar los valores $0, 1, 2, \ldots$, y en principio no ponemos una cota superior para el número de observaciones del evento. Adicionalmente supongamos que conocemos la tasa media de ocurrencia del evento de interés, que denotamos por la letra λ. El parámetro λ es positivo y se interpreta como el número promedio de ocurrencias del evento, por unidad de tiempo. La probabilidad de que la variable aleatoria X tome un valor x se definirá como se indica a continuación.

Decimos entonces que X tiene una distribución Poisson con parámetro $\lambda > 0$. Y escribimos $X \sim \text{Poisson}(\lambda)$ cuando

$$P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{si } x = 0, 1, 2, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

Puede demostrarse que cuando $X \sim \text{bin}(n, p)$ y hacemos tender n a infinito y p a cero de tal forma que el producto np se mantenga constante igual a λ, entonces la variable aleatoria X adquiere la distribución Poisson con parámetro λ.

Distribución Binomial Negativa

Si en una sucesión infinita de ensayos Bernoulli la variable aleatoria X cuenta el número de fracasos antes de obtener el r-ésimo éxito, entonces decimos que X tiene una distribución binomial negativa con parámetros p y r. Y escribimos $X \sim \text{bin neg}(p, r)$. En este caso tenemos que X puede tomar los valores $0, 1, 2, \ldots$ con probabilidades como se indica a continuación.

$$P(X = x) = \begin{cases} \begin{pmatrix} r + x - 1 \\ x \end{pmatrix} p^r (1-p)^x & \text{si } x = 0, 1, 2, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

Aparece el término p^r pues la sucesión de ensayos Bernoulli no concluye sino hasta obtener r éxitos. Podemos tener un número variable de fracasos, de ahi el término $(1 - p)^x$, y finalmente el factor $\begin{pmatrix} r + x - 1 \\ x \end{pmatrix}$ que nos dice las diferentes formas en que los r éxitos pueden aparecer en los $r + x - 1$ ensayos realizados antes del último que necesariamente fue un éxito.

Es claro que esta distribución es una generalización de la distribución geométrica.

Distribución Hipergeométrica

Supongamos que tenemos un conjunto de M objetos de los cuales K son de una primera clase y $M-K$ son de una segunda clase. Supongamos que de este conjunto tomamos una muestra aleatoria de tamaño n, la muestra es entonces sin reemplazo y el orden de los objetos seleccionados no importa. El espacio muestral de este experimento consiste entonces de todas las posibles muestras de tamaño n que se pueden obtener del conjunto mayor de tamaño M. La cardinalidad del espacio muestral es entonces $\binom{M}{n}$.

Si para cada muestra definimos la variable aleatoria X como el número de objetos de la primera clase contenidos en la muestra seleccionada, entonces X puede tomar los valores $0, 1, 2, \ldots, n$, suponiendo $n \leq K$. La probabilidad de que X tome un valor x estará dada por la fórmula que enunciamos a continuación.

Decimos que X tiene una distribución hipergeométrica com parámetros M, K y n. Y escribimos $X \sim \text{Hipergeo}(M, K, n)$ si

$$P(X = x) = \begin{cases} \dfrac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} & \text{si } x = 0, 1, 2, \ldots, n. \\ 0 & \text{otro caso.} \end{cases}$$

El término $\binom{K}{x}$ nos dice las diferentes formas en que de los K objetos de la primera clase se pueden escoger x de ellos y el término $\binom{M-K}{n-x}$ es nuevamente las diferentes formas de escoger $n-x$

objetos de la totalidad de $M - K$ objetos de la segunda clase. Usamos el principio multiplicativo para obtener el número total de muestras diferentes en donde x objetos son de la primera clase y $n - x$ objetos son de la segunda clase.

EJERCICIOS

1. Sea $X \sim \text{bin}(n, p)$. Definamos la v.a. $Y = n - X$. Demuestre que $Y \sim \text{bin}(n, 1 - p)$.

2. Sea $X \sim \text{bin}(n, p)$. Demuestre la siguiente propiedad recursiva
$$P(X = i + 1) = \left(\frac{p}{1 - p}\right)\left(\frac{n - i}{i + 1}\right) P(X = i).$$

3. Sea $X \sim \text{bin}(n, p)$. Encuentre el valor de x tal que $P(X = x)$ es máximo.

4. Sea $X \sim \text{Poisson}(\lambda)$. Demuestre la siguiente propiedad recursiva
$$P(X = i + 1) = \left(\frac{\lambda}{i + 1}\right) P(X = i).$$

5. Sea $X \sim \text{Poisson}(\lambda)$. Demuestre que
$$P(\text{“X es par”}) = \frac{1}{2}(1 + e^{-2\lambda}).$$

6. Sea $X \sim \text{Geo}(p)$. Demuestre que para $i, j = 0, 1, 2, \ldots$ se cumple la siguiente igualdad.
$$P(X \geq i + j | X \geq i) = P(X \geq j).$$

EJERCICIOS DE APLICACIÓN

1. Según las estadísticas de una compañia de seguros, la probabilidad de que una persona de una cierta población de 1500 asegurados, fallezca durante el transcurso del siguiente año es de 0.001.

   (a) ¿ Cuál es la probabilidad de que ocurran mas de cinco siniestros durante el siguiente año ?

   (b) ¿ Cuántos fallecimientos en promedio espera tener la compañia aseguradora durante el siguiente año ?

2. Una urna contiene cuatro canicas rojas y cinco azules. Se extrae al azar una muestra con reemplazo de tres canicas. Defina la variable aleatoria X como el número de canicas rojas en la muestra. Obtenga y grafique la función de densidad de X.

3. Suponga que se realizan n lanzamientos independientes de una moneda cuya probabilidad de caer "águila" es p. Demuestre que la probabilidad de que se obtenga un número par de "águilas" es $[(1-2p)^n - 1]/2$.

3.4 Distribuciones Continuas

Las siguientes son algunas distribuciones de probabilidad de variables aleatorias continuas.

Distribución Uniforme Continua

Decimos que una variable aleatoria X tiene una distribución uniforme continua en el intervalo $[a, b]$, y escribimos $X \sim \text{Unif}[a, b]$, cuando la función de densidad de X esta dada por

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in [a,b], \\ 0 & \text{otro caso.} \end{cases}$$

Distribución Exponencial

Decimos que una variable aleatoria continua X tiene una distribución exponencial con parámetro $\lambda > 0$, y escribimos $X \sim \exp(\lambda)$, cuando

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{otro caso.} \end{cases}$$

Distribución Gama

La variable aleatoria continua X tiene una distribución gama con parámetros $\lambda > 0$ y $z > 0$, y escribimos $X \sim \text{gama}(\lambda, z)$, si la función de densidad de X esta dada por

$$f_X(x) = \begin{cases} \dfrac{\lambda}{\Gamma(z)}(\lambda x)^{z-1} e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{otro caso.} \end{cases}$$

En la expresión anterior aparece el término $\Gamma(z)$. Esta es la **función gama** que se define como sigue

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt,$$

para valores de z para los cuales la integral es convergente. Esta función satisface las siguientes propiedades

1. $\Gamma(z+1) = z\Gamma(z)$.

2. $\Gamma(z+1) = z!$ si z es entero.

3. $\Gamma(2) = \Gamma(1) = 1$.

4. $\Gamma(1/2) = \sqrt{\pi}$.

5. Si $z = 1, 2, 3, \ldots$, entonces

$$\Gamma(z+1/2) = \frac{1 \cdot 3 \cdot 5 \cdots (2z-1)}{2^z}\sqrt{\pi}.$$

6. $\Gamma(z)\Gamma(1-z) = \dfrac{\pi}{\operatorname{sen}\, \pi z}$.

7. Fórmula de duplicación:

$$\Gamma(2z) = 2^{2z-1}\pi^{-1/2}\Gamma(z)\Gamma(z-1).$$

El lector interesado puede encontrar mayor información acerca de la función gama en [1]. El nombre de esta distribución de probabilidad es evidente. Observamos además que la distribución exponencial es un caso particular de la distribución gama. Si en la distribución gama tomamos el parámtro z igual a 1, lo que obtenemos es la distribución exponencial.

Distribución Beta

Decimos que la variable aleatoria continua X tiene una distribución beta con parámetros $a > 0$ y $b > 0$, y escribimos $X \sim \text{beta}(a, b)$, cuando

$$f_X(x) = \begin{cases} \dfrac{1}{B(a,b)}x^{a-1}(1-x)^{b-1} & \text{si } x \in (0,1), \\ 0 & \text{otro caso.} \end{cases}$$

El término $B(a, b)$ se conoce como la **función beta** y de alli es donde adquiere el nombre esta distribución de probabilidad. La función beta se define como sigue

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx.$$

Esta función satisface las siguientes propiedades

1. $B(a,b) = B(b,a)$.
2. $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Existen muchas otras distribuciones de probabilidad que se han utilizado en diferentes aplicaciones. Dentro de la lista anterior hemos omitido la más importante de todas, la distribución normal, pues dedicaremos una sección entera para estudiar con cierta detalle esta distribución.

EJERCICIOS

1. Demuestre que una variable aleatoria exponencial X con parámetro λ satisface la siguiente propiedad: Para todo $s,t \in (0,\infty)$
$$P(X > t+s | X > t) = P(X > s).$$
2. Sea X una variable aleatoria continua que toma valores en el intervalo $(0,\infty)$. Suponga que para cualesquiera dos números x_0 y x_1 en $(0,\infty)$ se cumple que
$$P(x_0 \leq X \leq x_0 + x_1 | X \geq x_0) = P(X \leq x_1).$$
Demuestre entonces que $P(X \leq x_1) = 1 - e^{-cx_1}$ para alguna constante positiva c.
3. Demuestre que
$$\Gamma(z) = \lim_{n\to\infty} n^z \frac{1\cdot 2\cdot 3\cdots n}{z(z+1)(z+2)\cdots(z+n)},$$
siguiendo los siguientes pasos.

(a) Defina
$$\Gamma(z,n) = \int_0^n (1-\frac{t}{n})^n t^{z-1}\,dt,$$
y observe que
$$\lim_{n\to\infty} \Gamma(z,n) = \Gamma(z).$$

(b) Realice el cambio de variable $u = t/n$ para demostrar que
$$\Gamma(z,n) = n^z \int_0^1 (1-u)^n u^{z-1}\,du.$$

(c) Aplique integración por partes n veces para encontrar que
$$\Gamma(z,n) = n^z \frac{n!}{z(z+1)\cdots(z+n-1)} \int_0^1 u^{z+n-1}\,du.$$

(d) Calcule la última integral y demuestre que
$$\Gamma(z,n) = n^z \frac{n!}{z(z+1)\cdots(z+n)}.$$
Finalmente tome el límite cuando n tiende a infinito y use el inciso (a).

4. Sea X una variable aleatoria continua con función de densidad dada por
$$f_X(x) = \begin{cases} cx^2 & \text{si } 0 \leq x \leq 1, \\ 0 & \text{otro caso.} \end{cases}$$

(a) Encuentre el valor de c y grafique $f_X(x)$.

(b) Encuentre valores de α y β tales que $P(\alpha \leq X \leq \beta) = 1/2$ y la diferencia $\beta - \alpha$ es mínima.

(c) Encuentre valores de α y β tales que $P(\alpha \leq X \leq \beta) = 1/4$ y la diferencia $\beta - \alpha$ es máxima.

3.5 La Distribución Normal

Dedicaremos una sección entera a la distribución normal pues es la distribución de probabilidad de mayor importancia.

Definición. *Decimos que la v.a. continua X tiene una* **distribución normal** *si su función de densidad $f_X(x)$ esta dada por*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-(x-\mu)^2/2\sigma^2\right],$$

en donde $\mu \in \mathbf{R}$ y $\sigma^2 > 0$ son dos parámetros. Y escribimos entonces $X \sim N(\mu, \sigma^2)$.

En particular, decimos que la v.a. X tiene una distribución **normal estándar** si tiene una distribución normal con parámetros $\mu = 0$ y $\sigma^2 = 1$. En este caso la función de densidad se reduce a la expresión

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-x^2/2\right].$$

Es posible transformar una variable aleatoria normal no estándar en una estándar mediante una operación sencilla.

Proposición. *Sea X una variable aleatoria con distribución normal con parámetros μ y σ^2. Entonces la variable aleatoria*

$$Z = \frac{X-\mu}{\sigma}$$

tiene una distribución normal estándar.

Demostración. Encontraremos primero $F_Z(z)$ y después derivaremos respecto de z para encontrar $f_Z(z)$.

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(\frac{X-\mu}{\sigma} \leq z) \end{aligned}$$

$$
\begin{aligned}
&= P(X \le \mu + \sigma z) \\
&= F_X(\mu + \sigma z)
\end{aligned}
$$

Por lo tanto

$$
\begin{aligned}
f_Z(z) &= \frac{d}{dz} F_Z(z) \\
&= \frac{d}{dz} F_X(\mu + \sigma z) \\
&= f_X(\mu + \sigma z)\, \sigma \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-x^2/2\right].
\end{aligned}
$$

•

Es común usar la letra Z para denotar una variable aleatoria con distribución normal estándar.

EJERCICIOS TEÓRICOS

1. Demuestre que la función de densidad normal con parámetros μ y σ^2 tiene un máximo absoluto en $x = \mu$, y tiene puntos de inflexión en $x = \mu \pm \sigma$.

2. Demuestre que para todo $x > 0$ se cumplen las desigualdades siguientes.

$$
\frac{x}{1+x^2} e^{-x^2/2} \le \int_x^{\infty} e^{-u^2/2}\, du \le \frac{1}{x} e^{-x^2/2}.
$$

EJERCICIOS DE APLICACIÓN

1. El tiempo de vida útil T de un televisor es una variable aleatoria con distribución normal con parámetros $\mu = 15$ años y $\sigma = 2$.

(a) ¿ Cuál es la probabilidad de que un televisor dure mas de 16 años ?

(b) ¿ Cuál es la probabilidad de que un televisor falle durante los primeros 10 años de vida ?

(c) Un televisor ha estado funcionando bien durante sus primeros diez años de vida, ¿ Cuál es la probabilidad de que sobrepase los 15 años de vida útil ?

2. Suponga que la calificación final de un alumno es una variable aleatoria normal con parámetros μ y σ^2. El profesor decide dar la calificación de 10 a aquellos estudiantes cuyo total de puntos excede el valor $\mu + 2\sigma$, 9 a aquellos cuyo puntaje está entre $\mu + \sigma$ y $\mu + 2\sigma$, 8 a aquellos con puntaje entre μ y $\mu + \sigma$, 7 a aquellos con puntaje entre $\mu - \sigma$ y μ, 6 a aquellos con puntaje entre $\mu - 2\sigma$ y $\mu - \sigma$, y 5 a aquellos con puntaje menor a $\mu - 2\sigma$. ¿ Qué porcentaje de alumnos recibirá cada calificación ? Compruebe que la suma de los porcentajes es 100%.

3. Una embotelladora de refrescos llena envases de 300 ml. Debido a imprecisiones de la máquina, la cantidad de refresco depositada en cada envase es variable pero puede modelarse con una variable aleatoria normal con parámetros $\mu = 300ml.$ y $\sigma = 10$.

   (a) ¿ Cuál es la probabilidad de llenar un envase con al menos 290 ml. de refresco ?

   (b) Según el estándar de calidad de la empresa, ningún envase debe contener menos de 270 ml. ¿ Cuantos refrescos de un lote de 1000 no pasarán este control de calidad ?

   (c) Si se utilizan envases de 320 ml., ¿ Cuántos envases se derramaràn en 1000 llenados ?

4. El diámetro de una cable de cobre en un proceso de manufactura puede modelarse mediante una variable aleatoria normal con parámetros $\mu = 0.20$ cm. y σ^2 con un valor desconocido.

¿ Qué valor debe tener σ^2 para que la probabilidad de que el diámetro diste de la media en al menos 0.1 cm. sea menor que 0.04 ?

Capítulo 4

Esperanza, Varianza, Momentos

En este capítulo estudiaremos algunas caracteristicas numéricas de variables aleatorias. Estas caracteristicas son por ejemplo el valor promedio que toma una v.a. o el grado de dispesión de los valores que puede tomar una v.a.

4.1 Esperanza

Una de las características numéricas de las variables aleatorias es la esperanza.

¿ Qué es la esperanza de una v.a. ?

Es un número real que denotaremos por $E(X)$ y que calcularemos para cada v.a. X. La definición formal es la siguiente.

Definición. *Sea X una v.a. continua con función de densidad $f_X(x)$, definimos la* **esperanza** *de X como sigue*

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

Analogamente, si X es una v.a. discreta con función de densidad $f_X(x)$, definimos la esperanza de X como sigue

$$E(X) = \sum_x x f_X(x),$$

en donde la suma se realiza sobre todos los valores x que la v.a. X puede tomar.

La esperanza se conoce también como **media**, **valor promedio**, o **valor esperado** de la v.a. Para algunas variables aleatorias la esperanza puede ser infinita y en ese caso decimos que la esperanza no existe o que la v.a. no tiene esperanza.

Ejemplo. Sea X una variable aleatoria discreta con función de densidad dada por la siguiente tabla

x	-1	0	1	2
$f_X(x)$	1/10	4/10	3/10	2/10

Entonces calculamos la esperanza de X como sigue

$$\begin{aligned} E(X) &= \sum_x x f_X(x) \\ &= -1(1/10) + 0(4/10) + 1(3/10) + 2(2/10) \\ &= 6/10. \end{aligned}$$

El ejemplo anterior nos ilustra el hecho de que el número $E(X)$ no necesariamente es un valor de los que puede tomar la variable aleatoria X. En nuestro ejemplo, el valor $6/10$ no figura dentro de los posibles valores de la variable aleatoria.

Ejemplo. Sea X una variable aleatoria continua con función de densidad dada por

$$f_X(x) = \begin{cases} 4x & \text{si } x \in [0, 1/\sqrt{2}], \\ 0 & \text{otro caso.} \end{cases}$$

Entonces por definición

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x)\, dx \\ &= \int_0^{1/\sqrt{2}} x(4x)\, dx \\ &= \frac{4}{3}x^3 \Big|_0^{1/\sqrt{2}} \\ &= \frac{\sqrt{2}}{3}. \end{aligned}$$

De esta forma vemos que el cálculo de una esperanza se reduce a calcular correctamente una suma o una integral.

¿ Para qué sirve la esperanza ?

La esperanza se puede interpretar como el valor promedio de la variable aleatoria. Efectivamente nuestra definición de esperanza no es sino un promedio ponderado de los valores que toma la v.a. Conocer $E(X)$ para una v.a. X nos proporciona información acerca de cual es el valor central que puede tomar la v.a. X. Por ejemplo, si $E(X) = 5/2$ entonces sabemos que los valores que X puede tomar se centran (ponderadamente) alrededor del número $5/2$.

¿ Qué propiedades tiene la esperanza ?

Enunciamos a continuación algunas propiedades de la esperanza, la mayoría de las cuales se siguen directamente de la definición. Supondremos que tanto X como Y son variables aleatorias con esperanza finita.

Proposición. *Si c es una constante entonces $E(c) = c$.*

Demostración. Como $X = c$ es una variable aleatoria discreta que toma únicamente un valor, tenemos que $P(X = c) = 1$ y entonces

$$E(X) = c\,P(X = c) = c.$$

•

Proposición. $E[E(X)] = E(X)$.

Demostración. Esta es una simple consecuencia de la proposición anterior. Sabemos que $E(X)$ es un número, por lo tanto su esperanza es ese mismo número. •

Proposición. *Si $X \geq 0$ entonces $E(X) \geq 0$.*

Demostración. Suponiendo el caso continuo tenemos que como $X \geq 0$ entonces $f_X(x) = 0$ para $x < 0$ y entonces

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx \\ &= \int_{0}^{\infty} x f_X(x)\,dx \\ &\geq 0. \end{aligned}$$

•

Proposición. *Si c es una constante entonces*

$$E(cX) = cE(X).$$

Demostración. Probaremos esta propiedad en el caso discreto siendo la demostración análoga para el caso continuo. Si X es entonces una variable aleatoria discreta que toma los valores $x_0, x_1, \ldots$ con probabilidades respectivas $P(X = x_0), P(X = x_1), \ldots$, entonces la variable aleatoria cX toma los valores $cx_0, cx_1, \ldots$ con las mismas probabilidades que antes. Entonces tenemos que

$$\begin{aligned} E(cX) &= \sum_i cx_i P(X = x_i) \\ &= c \sum_i x_i P(X = x_i) \\ &= cE(X). \end{aligned}$$

•

Proposición. *E(X+Y)=E(X)+E(Y).*

Demostración. Consideremos nuevamente el caso de variables aleatorias discretas. Supongamos que $x_0, x_1, \ldots$ son valores que puede tomar X y que $y_0, y_1, \ldots$ son los correspondientes valores para Y. Entonces

$$\begin{aligned} E(X+Y) &= \sum_{i,j} (x_i + y_j) P(X = x_i, Y = y_j) \\ &= \sum_{i,j} x_i P(X = x_i, Y = y_j) \\ &\qquad + \sum_{i,j} y_j P(X = x_i, Y = y_j) \\ &= \sum_i x_i P(X = x_i) + \sum_j y_j P(Y = y_j) \\ &= E(X) + E(Y). \end{aligned}$$

Observemos que hemos utilizado el teorema de probabilidad total en nuestra demostración al reducir las dobles sumas en sumas respecto de un solo índice. •

Observamos tambien que las anteriores dos propiedades nos dicen que la esperanza es lineal, lo cual resulta ser de mucha utilidad.

Proposición. *Si* $X \geq Y$ *entonces* $E(X) \geq E(Y)$.

Demostración. Por hipótesis sabemos que $X - Y \geq 0$. Entonces $E(X - Y) \geq 0$. Por la propiedad anterior tenemos finalmente que $E(X) - E(Y) \geq 0$. •

Proposición. *Si* X *y* Y *son independientes entonces*

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

Demostración. Consideremos nuevamente el caso de variables aleatorias discretas. Supongamos que $x_0, x_1, \ldots$ son valores que puede tomar X y que $y_0, y_1, \ldots$ son los correspondientes valores para Y. Entonces como X y Y son variables aleatorias independientes tenemos que

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j).$$

Y por lo tanto

$$\begin{aligned} E(X \cdot Y) &= \sum_{i,j} (x_i \cdot y_j) P(X = x_i, Y = y_j) \\ &= \sum_{i,j} (x_i \cdot y_j) P(X = x_i) P(Y = y_j) \\ &= \sum_i x_i P(X = x_i) \ \sum_j y_j P(Y = y_j) \\ &= E(X)\ E(Y). \end{aligned}$$

•

EJERCICIOS

1. Decimos que una variable aleatoria X tiene función de densidad f_X simétrica respecto del número a si $f_X(a - x) =$

$f_X(a+x)$ para todo número real x. Suponiendo lo anterior, demuestre entonces que $E(X) = a$ cuando la esperanza existe. Demuestre además que $E[(X-a)^n] = 0$ para n un entero impar positivo, suponiendo que $E(X^n)$ existe.

2. Sea X una variable aleatoria discreta con función de densidad

$$f_X(x) = \begin{cases} \dfrac{1}{x(x+1)} & \text{si } x = 1, 2, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

   (a) Demuestre que f_X es función de densidad.

   (b) Demuestre que $E(X)$ no existe.

3. Demuestre que si X es una variable aleatoria no negativa entonces

   (a) $E(X) = \displaystyle\sum_{x=0}^{\infty} P(X > x)$ si X es una v.a. discreta.

   (b) $E(X) = \displaystyle\int_0^{\infty} P(X > x)$ si X es una v.a. continua.

   Use este resultado para calcular $E(X)$ cuando $X \sim \text{Geo}(p)$.

4. Sea $X \sim \text{Poisson}(\lambda)$ tal que $P(X = 0) = P(X = 1)$. Demuestre que $E(X) = 1$.

5. Sea $X \sim \text{Unif}[1, 2]$. Encuentre el valor de θ que satisface la ecuación $P[X > \theta + E(X)] = 1/6$. Respuesta: $\theta = 1/3$.

6. Sea X una variable aleatoria continua con función de densidad dada por

$$f_X(x) = \begin{cases} a + bx & \text{si } 0 \leq x \leq 1 \\ 0 & \text{otro caso.} \end{cases}$$

   Suponga que $E(X) = 1$. Demuestre entonces que las constantes a y b son tales que $a = 2$ y $b = -2$.

4.2 Esperanza de Funciones de Variables Aleatorias

Muy frecuentemente es necesario calcular la esperanza de una función de una variable aleatoria, por ejemplo $E(X^3)$ o $E(2e^X)$. El siguiente resultado nos dice cómo llevar a cabo este tipo de cálculos de una manera más sencilla a la forma de aplicar directamente nuestra definición de esperanza.

Teorema
Sea X una variable aleatoria continua con función de densidad f_X. Sea $g : R \to R$ una función cualquiera tal que $g(X)$ es una variable aleatoria. Si la esperanza de la v.a. $g(X)$ existe, se puede calcular como sigue

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx.$$

Demostración. Supondremos el caso de variables aleatorias discretas. Sean $\text{Rango}(X) = \{x_1, x_2, \ldots\}$ y $\text{Rango}(Y) = \{y_1, y_2, \ldots\}$, el conjunto de valores que toman las variables aleatorias X y Y respectivamente. Estos conjuntos pueden ser finitos o infinitos.

Sucede el evento $(Y = y_i)$ si y solo si existe por lo menos un valor $x \in \{x_1, x_2, \ldots\}$ tal que $y_i = g(x)$. Agrupémos todos esos valores x en el conjunto A_i como definimos a continuación.

$$A_i = \{x \in \text{Rango}(X) : g(x) = y_i\}.$$

A los elementos de A_i los enumeramos $x_i^1, x_i^2, \ldots$ y entonces tenemos que

$$P(Y = y_i) = P(X \in A_i) = \sum_j P(X = x_i^j).$$

Por lo tanto

$$\begin{aligned}
E(Y) &= \sum_i y_i P(Y = y_i) \\
&= \sum_i y_i P(X \in A_i) \\
&= \sum_i \sum_j g(x_i^j) P(X = x_i^j) \\
&= \sum_k g(x_k) P(X = x_k).
\end{aligned}$$

•

Hemos enunciado el resultado anterior para variables aleatorias continuas pero hemos hecho la demostración para variables aleatorias discretas. En ambos casos nuestro resultado es válido.

¿ Por qué es útil el teorema anterior ?

La enorme ventaja del teorema anterior radica en lo siguiente. Si queremos calcular la esperanza de, por ejemplo, la v.a. X^3, para hacer este cálculo usando nuestra definición de esperanza tendríamos que conocer la función de densidad de la v.a. X^3, esto es f_{X^3}, lo cual puede a veces ser muy complicado, y calcular entonces

$$E(X^3) = \int_{-\infty}^{\infty} x f_{X^3}(x)\, dx.$$

El teorema anterior establece que no es necesario conocer la función de densidad de X^3, basta con saber la función de densidad de X y hacer los cálculos como indica la fórmula, esto es

$$E(X^3) = \int_{-\infty}^{\infty} x^3 f_X(x)\, dx.$$

Ejemplo. Sea X una variable aleatoria continua uniforme en el intervalo $[0, \pi]$. Y sea $Y = \sin X$. Encuentre $E(Y)$.

Solución. Por el teorema anterior,

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} \sin x\, f_X(x)\, dx \\ &= \int_0^{\pi} \frac{1}{2\pi} \sin x\, dx \\ &= \frac{1}{\pi}. \end{aligned}$$

EJERCICIOS

1. Sea $X \sim \mathrm{N}(0, \sigma^2)$. Encuentre la esperanza y la varianza de la variable aleatoria $Y = |X|$.

2. Sea $X \sim \mathrm{Unif}[0,1]$. Encuentre la esperanza y la varianza de la variable aleatoria $Y = \cos \pi X$. Encuentre además el n-ésimo momento de Y.

4.3 Varianza

Una segunda característica numérica de la variables aleatoria se conoce con el nombre de varianza.

¿ Qué es la varianza de una v.a. ?

Es un número real no negativo que denotaremos por $\mathrm{Var}(X)$ y que calcularemos como establece nuestra siguiente definición.

Definición. *Sea X una variable aleatoria con esperanza finita $E(X)$. Definimos la* **varianza** *de X como el número*

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

Cuando X es una v.a. continua con función de densidad $f_X(x)$, nuestra definición nos dice que calcularemos la varianza de X como sigue

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x)\, dx.$$

Mientras que cuando X es una v.a. discreta con valores $x_0, x_1, x_2, \ldots$ entonces

$$\mathrm{Var}(X) = \sum_{i=0}^{\infty} (x_i - E(X))^2 f_X(x_i).$$

Debemos enfatizar que la varianza es siempre un número no negativo. Existen variables aleatorias con varianza infinita.

¿ Qué interpretación tiene la varianza ?

La varianza mide el grado de dispersión de los valores que toma la v.a. respecto a la media $E(X)$. De manera informal e imprecisa, podemos decir que cuando una variable aleatoria tiene una varianza grande significa que los valores que toma la variable aleatoria estan, en promedio, muy dispersos. Mientras que una varianza pequeña significa que los valores se encuentran cercanos, en promedio, a la media $E(X)$.

¿ Qué propiedades tiene la varianza ?

Enunciaremos y probaremos a continuación una serie de propiedades que satisface la varianza. En las proposiciones siguientes supondremos que X y Y son variables aleatorias con varianza finita.

Proposición. *Si c es una constante entonces $Var(c) = 0$.*

Demostración. La variable aleatoria constante $X = c$ es una v.a. discreta que toma un solo valor, el valor c, con probabilidad 1. Por lo tanto $E(X) = c$ y entonces

$$\text{Var}(X) = (c - c)^2 = 0.$$

•

Proposición. *$Var(X) = E(X^2) - E^2(X)$.*

Demostración. Usaremos aqui las propiedades de la esperanza. Empezamos escribiendo la definición de varianza.

$$\begin{aligned} \text{Var}(X) &= E\left[(X - E(X))^2\right] \\ &= E\left[X^2 - 2XE(X) + E^2(X)\right] \\ &= E(X^2) - 2E(X)E(X) + E^2(X) \\ &= E(X^2) - E^2(X). \end{aligned}$$

•

Proposición. *Si c es una constante entonces*

$$Var(cX) = c^2\, Var(X).$$

Demostración. Por la propiedad anterior tenemos que

$$\begin{aligned} \text{Var}(cX) &= E(c^2X^2) - E^2(cX) \\ &= c^2E(X^2) - c^2E^2(X) \\ &= c^2\text{Var}(X). \end{aligned}$$

•

En particular esta propiedad dice que la varianza no es lineal pues las constantes se elevan al cuadrado al salir de la varianza.

Proposición. *Si X y Y son independientes entonces*

$$Var(X + Y) = Var(X) + Var(Y).$$

Demostración.

$$
\begin{aligned}
\mathrm{Var}(X+Y) &= E\left[((X+Y)-E(X+Y))^2\right] \\
&= E\left[\{(X-E(X))-(Y-E(Y))\}^2\right] \\
&= E\left[(X-E(X))^2\right]-2E\left[(X-E(X))(Y-E(Y))\right] \\
&\quad +E\left[(Y-E(Y))^2\right] \\
&= \mathrm{Var}(X)+\mathrm{Var}(Y)-2E\left[(X-E(X))(Y-E(Y))\right].
\end{aligned}
$$

Debido a la independencia entre X y Y, el último sumando se anula. Esto concluye nuestra demostración. •

EJERCICIOS

1. Demuestre que para todo $c \in R$,

$$E[(X-c)^2] = Var(X) + (E(X)-c)^2.$$

2. Sea $g(c) = E(X-c)^2$. Demuestre que la función g tiene un mínimo en $c_0 = E(X)$ y que $g(c_0) = Var(X)$.

3. Sea X una variable aleatoria cualquiera cuyos posibles valores estan contenidos en el intervalo $[a,b]$. Demuestre que

   (a) $a \leq E(X) \leq b$.

   (b) $0 \leq \mathrm{Var}(X) \leq (b-a)^2/4$.

4. Sea X una variable aleatoria distribuida normalmente con media $\mu = 0$ y varianza $\sigma^2 > 0$. Sea $Y = |X|$. Encuentre la media y la varianza de Y.

5. Sean a y b dos números reales tales que $a < b$. Dividimos el intervalo $[a,b]$ en n partes iguales y consideremos los $(n+1)$ valores de la partición

$$
\begin{aligned}
x_0 &= a \\
x_1 &= a+(b-a)/n
\end{aligned}
$$

$$\begin{aligned} x_2 &= a + 2(b-a)/n \\ &\vdots \\ x_n &= b. \end{aligned}$$

Sea X una variable aleatoria discreta que toma cada uno de estos valores con idéntica probabilidad $1/(n+1)$.

(a) Demuestre que $E(X) = (a+b)/2$.

(b) Calcule $\mathrm{Var}(X)$ y demuestre que

$$\lim_{n\to\infty} \mathrm{Var}(X) = (b-a)/12.$$

4.4 Momentos

Definiremos en esta sección ciertas características numéricas de variables aleatorias que llamaremos **momentos**. Estas caracteristicas numéricas que definiremos a continuación incluyen a la esperanza y la varianza mencionadas anteriormente como casos particulares.

Definición. *Sea X una variable aleatoria con media μ, y sea n un número natural. Si existe, definimos*

1. *el n-ésimo momento como $E(X^n)$.*
2. *el n-ésimo momento central como $E((X-\mu)^n)$.*
3. *el n-ésimo momento absoluto como $E(|X|^n)$.*
4. *el n-ésimo momento central absoluto como $E(|X-\mu|^n)$.*
5. *el n-ésimo momento generalizado respecto a c como $E((X-c)^n)$.*

Observamos entonces que el primer momento de X es simplemente el la esperanza $E(X)$. Y el segundo momento central de X es la varianza $\mathrm{Var}(X)$.

¿ Para qué sirven los momentos ?

Hemos interpretado el primer momento, $E(X)$, como el valor promedio ponderado de los valores que toma la variable aleatoria X. Y el segundo momento central, esto es la varianza $\mathrm{Var}(X)$, como una medida de dispersión del conjunto de valores tomado por la variable aleatoria X. De modo que si conocemos la esperanza y la varianza de una variable aleatoria tenemos cierta información de esa variable aleatoria. En general y bajo ciertas condiciones, los momentos caracterizan de manera única a una distribución de probabilidad. Enunciaremos ahora dos resultados al respecto.

Criterio de Carleman
Si X es una variable aleatoria tal que $E(X^n) < \infty$ para $n = 1, 2, \ldots$ y

$$\sum_{n=1}^{\infty} \left[E(X^{2n})\right]^{1/2n} = \infty,$$

entonces la sucesión de momentos $E(X), E(X^2), E(X^3), \ldots$ determina de manera única a la función de distribución $F_X(x)$.

Criterio de Hausdorff
Si X es una variable aleatoria acotada (es decir, existe una constante M tal que $P(|X| < M) = 1$) entonces la sucesión de momentos $E(X), E(X^2), E(X^3), \ldots$ determina de manera única a la función de distribución $F_X(x)$.

Por lo tanto, bajo alguna de las condiciones arriba mencionadas, conocer los momentos de X es equivalente a conocer F_X. Y con ello conocemos todo acerca de la variable aleatoria X.

EJERCICIOS

1. Calcule el n-ésimo momento de una variable aleatoria uniforme continua en el intervalo $[a, b]$.
2. Demuestre que si $E(X^2) < \infty$ entonces $E(X) < \infty$.
3. Demuestre que si $E(X^n) < \infty$ entonces $E(X^j) < \infty$ para $j = 0, 1, 2, \ldots, n$.
4. Sea X una variable aleatoria acotada, es decir, existe una constante M tal que

$$P(\omega \in \Omega : |X(\omega)| < M) = 1.$$

Demuestre entonces que $E(X^n)$ existe para todo número natural n.

4.5 La Función Generadora de Momentos

Definición. *Sea X una variable aleatoria. Definimos la* **función generadora de momentos** *(fmg) de la v.a. X, denotada por $m_X(t) : R \to R$, como sigue*

$$m_X(t) = E(e^{tX}),$$

cuando esta esperanza exista para ciertos valores de $t \in \mathbf{R}$.

Existen variables aleatorias para las cuales no existe una función generadora de momentos asociada. El nombre de esta función proviene de la siguiente propiedad importante.

Proposición. *Sea* X *una v.a. con fgm* $m_X(t)$ *tal que* $E(X^n)$ *existe para alguna* $n = 1, 2, 3, \ldots$. *Entonces*

$$\frac{d^n}{dt^n} m_X(t)\bigg|_{t=0} = E(X^n).$$

Demostración.

$$\frac{d^n}{dt^n} m_X(t) = \frac{d^n}{dt^n} E(e^{tX}) = E\left(\frac{d^n}{dt^n} e^{tX}\right) = E(X^n e^{tX}).$$

La evaluación en $t = 0$ nos da el resultado requerido. •

Finalmente enunciaremos un resultado que resulta ser una herramienta poderosa para encontrar la distribución de probabilidad de una v.a., especialmente cuando ésta es la suma de otras variables aleatorias independientes.

Teorema
Dos variables aleatorias tienen la misma distribución de probabilidad si y solo si sus correspondientes funciones generadoras de momentos coinciden.

Es decir, existe una relación biunívoca entre las distribuciones de probabilidad y las funciones generadoras de momentos. He aqui algunas propiedades de toda función generadora de momentos.

Proposición 1 *Sean* X *y* Y *dos v.a.s con fgm* $m_X(t)$ *y* $m_Y(t)$ *respectivamente, y sea* a *una constante. Entonces*

1. $m_{aX}(t) = m_X(at)$ *en donde* a *en una constante.*
2. $m_{X+a} = e^{at} m_X(t)$.
3. $m_{X+Y}(t) = m_X(t) \cdot m_Y(t)$ *cuando* X *y* Y *son independientes.*

Estas propiedades se demuestran directamente de la definición de fgm. La última propiedad puede sin dificultad extenderse para la suma finita de variables aleatorias independientes.

EJERCICIOS

1. Sea X una variable aleatoria discreta con función de densidad

$$f_X(x) = \begin{cases} k(1/5)^x & \text{si } x = 0, 1, 2, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

   En donde k es una constante positiva.

   (a) Encuentre el valor de la constante k.

   (b) Calcule $m_X(t)$.

   (c) Use $m_X(t)$ para calcular $E(X)$.

   (d) Use $m_X(t)$ para calcular $E(X^2)$ y $\text{Var}(X)$.

2. Sea X una variable aleatoria con fgm $m_X(t)$. Defina la función

$$R_X(t) = \ln m_X(t).$$

   Demuestre que

   (a) $R_X^{'}(t) = E(X)$.

   (b) $R_X^{''}(t) = \text{V}ar(X)$.

   (c) Use los incisos anteriores para calcular la media y la varianza de una v.a. con fgm dada por $m_X(t) = \exp[2(e^{2t} - 5)]$.

3. Sea X una variable aleatoria con función de densidad dada por

$$f_X(x) = \frac{1}{2}\exp(-|x|) \qquad \text{para} \quad -\infty < x < \infty.$$

   (a) Grafique $f_X(x)$.

   (b) Compruebe que $f_X(x)$ es una función de densidad.

(c) Demuestre que $m_X(t) = 1/(1 - t^2)$.

4. Sea X una variable aleatoria tal que $m_X(t) = \exp(2t + 4t^2)$. Encuentre la fgm de la variable aleatoria

$$Y = (X - 2)/10.$$

Use $m_Y(t)$ para encontrar $E(Y)$ y $\mathrm{Var}(Y)$.

Capítulo 5

Teoremas Límite

5.1 Vectores Aleatorios

Esta sección contiene una breve introducción a las variables aleatorias multidimensionales o vectores aleatorios. Consideraremos únicamente vectores aleatorios de dimensión dos aunque todas las definiciones y resultados que mencionaremos se pueden extender fácilmente para vectores de cualquier dimensión mayor.

Definición. *Sea* $(\Omega, \mathcal{F}, P)$ *un espacio de probabilidad. Un* **vector aleatorio** *de dimensión* 2 *es una función*

$$(X, Y) : \Omega \to R^2,$$

dada por $(X,Y)(\omega) = (X(\omega), Y(\omega))$ *en donde* X *y* Y *son variables aleatorias.*

Nuevamente diremos que un vector aleatorio es **discreto** si las variables aleatorias X y Y son discretas. Y llamaremos al vector **continuo** cuando las variables aleatorias que lo conforman son continuas.

Definimos a continuación la función de densidad conjunta de dos variables aleatorias.

Definición. *Sea (X, Y) un vector aleatorio discreto que toma los valores en el conjunto $Ran(X, Y) = \{(x_1, y_1), \ldots, (x_n, y_m)\}$. Definimos la* **función de densidad conjunta** *de (X, Y), denotada por $f_{X,Y}(x, y) : R^2 \to R$, como sigue*

$$f_{X,Y}(x,y) = \begin{cases} P(X = x, Y = y) & \text{si } (x,y) \in Ran(X,Y), \\ 0 & \text{otro caso.} \end{cases}$$

Si (X, Y) es un vector aleatorio continuo, decimos que la función $f_{X,Y}(x, y)$ es la función de densidad conjunta de (X, Y) si para todo $(x, y) \in R^2$

$$P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, dv\, du.$$

Tenemos las siguientes dos propiedades que caracterizan a toda función de densidad.

Proposición. *Toda función de densidad $f_{X,Y}(x, y)$ satisface las siguientes dos propiedades*

1. $f_{X,Y}(x, y) \geq 0$.
2. $\displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1$.

Recíprocamente decimos que una función $f : R^2 \to R$ es de densidad si cumple con las dos condiciones arriba mencionadas.

Definición. *Sea $f_{X,Y}(x, y)$ la función de densidad del vector aleatorio continuo (X, Y). Definimos la función de* **densidad marginal**

de la v.a. X como

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy,$$

y la correspondiente función de **densidad marginal** *de la v.a. Y como*

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx.$$

Hemos definido las densidades marginales para vectores aleatorios continuos. La correspondiente definición para vectores discretos involucra una suma en lugar de la integral. Es fácil verificar que estas densidades marginales son efectivamente funciones de densidad.

Definimos ahora la función de distribución conjunta de dos variables aleatorias.

Definición. *Sea (X,Y) un vector aleatorio. Definimos la* **función de distribución** *de (X,Y), denotada por $F_{X,Y} : R^2 \to R$, como la siguiente probabilidad conjunta. Para todo $(x,y) \in R^2$,*

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y).$$

Enunciamos a continuación algunas propiedades de toda función de distribución conjunta.

Proposición. *Sea (X,Y) un vector aleatorio con función de distribución conjunta $F_{X,Y}$. Tenemos entonces las siguientes propiedades*

1. $\lim_{x,y\to\infty} F_{X,Y}(x,y) = 1$.
2. $\lim_{x\to-\infty} F_{X,Y}(x,y) = 0$. *Análogamente cuando es y quien tiende a menos infinito.*
3. $F_{X,Y}(x,y)$ *es continua por la derecha en cada variable.*

4. *$F_{X,Y}(x,y)$ es una función monótona no decreciente en cada variable.*

Definición. *Decimos que una función bivariada $F(x,y) : \mathbf{R}^2 \to \mathbf{R}$ es una función de distribución si satisface las siguientes cuatro propiedades.*

1. $\lim_{x,y\to\infty} F(x,y) = 1$.
2. $\lim_{x\to-\infty} F(x,y) = 0$. *Análogamente cuando y tiende a menos infinito.*
3. *$F(x,y)$ es continua por la derecha en cada variable.*
4. *$F(x,y)$ es una función monótona no decreciente en cada variable.*

Definición. *Sea $f_{X,Y}(x,y)$ la función de densidad del vector aleatorio continuo (X,Y). Definimos la función de* **distribución marginal** *de la v.a. X como*

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y),$$

y la correspondiente función de **distribución marginal** *de la v.a. Y como*

$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y).$$

¿ Cómo encontramos $F_{X,Y}(x,y)$ a partir de $f_{X,Y}(x,y)$?

Conociendo la función de densidad conjunta $f_{X,Y}(x,y)$, es posible encontrar al función de distribución conjunta $F_{X,Y}(x,y)$ simplemente integrando en el caso continuo o sumando en el caso discreto. Para el caso continuo tenemos

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du.$$

En el caso discreto se suman todos los valores de $f_{X,Y}(u,v)$ para valores de u menores o iguales a x y valores de v menores o iguales a y.

¿ Cómo encontramos $f_{X,Y}(x,y)$ a partir de $F_{X,Y}(x,y)$?

Como sabemos que $f_{X,Y}(x,y)$ y $F_{X,Y}(x,y)$ guardan la relación

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\, dv\, du,$$

por el teorema fundamental del cálculo tenemos que

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$$

EJERCICIOS

1. Sea (X,Y) un vector aleatorio discreto con función de densidad $f_{X,Y}(x,y)$ dada por la siguiente tabla.

$x \setminus y$	-1	0	1
2	2/20	3/20	3/20
3	1/20	1/20	2/20
4	2/20	5/20	1/20

   (a) Grafique y demuestre que $f_{X,Y}(x,y)$ es efectivamente una función de densidad bivariada.

   (b) Calcule $P(X \in [3,5], Y \in [0,2])$.

   (c) Calcule $P(X+Y \leq 3)$.

   (d) Calcule las densidades marginales $f_X(x)$ y $f_Y(y)$. Grafique estas funciones verificando que son funciones de densidad.

   (e) Calcule y grafique $F_{X,Y}(x,y)$.

(f) Calcule las distribuciones marginales $F_X(x)$ y $F_Y(y)$. Grafique estas funciones y verifique que son efectivamente funciones de distribución.

2. Sea (X, Y) un vector aleatorio continuo con función de densidad $f_{X,Y}(x, y)$ dada por

$$f_{X,Y}(x,y) = \begin{cases} kx^2y & \text{si } x \in [0,1], y \in [0,1], \\ 0 & \text{otro caso.} \end{cases}$$

(a) Encuentre el valor de la constante k y grafique $f_{X,Y}(x, y)$.

(b) Calcule $P(X \in [1/2, 2], Y \in [0, 1/4])$.

(c) Calcule las densidades marginales $f_X(x)$ y $f_Y(y)$. Grafique estas funciones verificando que son funciones de densidad.

(d) Calcule y grafique $F_{X,Y}(x, y)$.

(e) Calcule las distribuciones marginales $F_X(x)$ y $F_Y(y)$. Grafique estas funciones y verifique que son efectivamente funciones de distribución.

5.2 La Desigualdad de Chebyshev

El siguiente resultado es bastante conocido y es de carácter teórico.

Desigualdad de Chebyshev
Sea X una v.a. con media μ y varianza finita σ^2. Para cualquier número real $k > 0$ tenemos que

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2.$$

Demostración. Supongamos que X es una v.a. continua con función de densidad $f_X(x)$. Entonces

$$\begin{aligned}
\sigma^2 &= E\left[(X-\mu)^2\right] \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx \\
&\geq \int_{|x-\mu|\geq k\sigma} (x-\mu)^2 f_X(x)\,dx \\
&\geq k^2\sigma^2 \int_{|x-\mu|\geq k\sigma} f_X(x)\,dx \\
&= k^2\sigma^2 P(|x-\mu| \geq k\sigma).
\end{aligned}$$

Despejando $P(\cdot)$ se obtiene la desigualdad de Chebyshev. •

Enunciamos a continuación algunas otras versiones de esta desigualdad que se obtienen facilmente de la demostrada.

1. $P(|X-\mu| \geq \epsilon) \leq \sigma^2/\epsilon^2$.
2. $P(|X-\mu| \leq k\sigma) \geq 1 - 1/k^2$.
3. $P(|X-\mu| < \epsilon) \geq 1 - \sigma^2/\epsilon^2$.

Estas versiones se obtienen de la demostrada tomando complementos o bien una elección apropiada de la constante k. Usaremos la desigualdad de Chebyshev para demostrar, en la siguiente sección, la ley débil de los grandes números.

EJERCICIOS

1. *La desigualdad de Markov.* Sea X una variable aleatoria no negativa con media finita μ. Demuestre que para cualquier constante positiva a,

$$P(X \geq a) \leq \mu/a.$$

Sugerencia: Utilice la misma técnica que la que usamos para demostrar la desigualdad de Chebyshev. Esta vez empiece escribiendo la definición de esperanza para una v.a. X continua.

2. Use la desigualdad de Markov para demostrar la desigualdad de Chebyshev. Sugerencia: Aplique la desigualdad de Markov a la variable aleatoria no negativa $(X-\mu)^2$. Esta es una demostración alternativa de la desigualdad de Chebyshev.

3. Sea X una variable aleatoria con media μ y varianza σ^2. Use la desigualdad de Chebyshev para estimar la probabilidad de que la v.a. X tome valores entre $\mu - k\sigma$ y $\mu + k\sigma$ para cualquier constante positiva k.

4. Sea X una variable aleatoria discreta con función de densidad dada por

$$f_X(x) = \begin{cases} 1/18 & \text{si } x = -1, 1 \\ 16/18 & \text{si } x = 0, \\ 0 & \text{otro caso.} \end{cases}$$

   (a) Calcule $E(X)$.
   (b) Calcule $\mathrm{Var}(X)$.
   (c) Calcule exactamente $P(|X-\mu| \geq 3\sigma)$.
   (d) Ahora use la desigualdad de Chebyshev para estimar $P(|X-\mu| \geq 3\sigma)$.

   Los cálculos de los incisos (c) y (d) deben coincidir. Este ejercicio demuestra que en general la cota superior dada por la desigualdad de Chebyshev es óptima, es decir, no puede establecerse una cota superior más pequeña.

5. Considere la versión de la desigualdad de Chebyshev

$$P(|X-\mu| < k\sigma) \geq 1 - 1/k^2.$$

Encuentre el mínimo valor de k de tal modo que la probabilidad de que una variable aleatoria tome valores entre $\mu - k\sigma$ y $\mu + k\sigma$ sea

(a) al menos 0.90.

(b) al menos 0.95.

5.3 La Ley de los Grandes Números

En esta sección estudiaremos uno de los teoremas más importantes de la teoría de la probabilidad clásica. Se le conoce como la ley de los grandes números. Existen varias versiones de este resultado, no es nuestro objetivo dar la versión más general sino comprender lo que dice. Para ello necesitamos mencionar algunos conceptos de convergencia de variables aleatorias.

Sea $X_1, X_2, X_3, \ldots$ una sucesión infinita de variables aleatorias. Estamos interesados en conocer cuál el comportamiento límite de esta sucesión. Si evaluamos cada una de las variables aleatorias X_i en un elemento ω del espacio muestral Ω, lo que se obtiene es un númeo real para cada índice i. Y entonces lo que se tiene es una sucesión bien definida de números reales $X_1(\omega), X_2(\omega), X_3(\omega), \ldots$. Y para esta sucesión es válido preguntarnos por la convergencia en $\mathbf{R}$. Si existe una variable aleatoria X tal que para casi toda ω, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$, entonces decimos que la sucesión de variables aleatorias converge a la v.a. X con probabilidad 1, A este tipo de convergencia también se le llama convergencia casi siempre. Daremos ahora la definición precisa de este tipo de convergencia.

Definición. *Decimos que la sucesión de variables aleatorias $X_1, X_2, \ldots$ converge* **casi siempre** *(o* **con probabilidad 1***) a la variable aleatoria X si*

$$P\left(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1.$$

Nuestra primera versión de la ley de los grandes números establece que, bajo ciertas condiciones, el promedio de variables aleatorias $(1/n)\sum_{i=1}^{n} X_i$ converge casi siempre a una constante.

La Ley Fuerte de los Grandes Números
Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes e identicamente distribuidas (v.a.i.i.d.) tales que $E(X_i) = \mu$. Entonces

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right) = 1.$$

En la teoría de la probabilidad existen otras formas en que una sucesión de variables aleatorias puede converger a una variable aleatoria. Tenemos por ejemplo la convergencia en probabilidad que definimos a continuación

Definición. *Decimos que la sucesión de variables aleatorias $X_1, X_2, \ldots$ converge* **en probabilidad** *a la variable aleatoria X si para toda $\epsilon > 0$ tenemos que*

$$\lim_{n\to\infty} P\left(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}\right) = 0.$$

De inmediato surgen preguntas acerca de la relación que existe entre este tipo de convergencia y la definida anteriormente. Diremos únicamente que la convergencia casi siempre implica la convergencia en probabilidad y que el recíproco es falso.

Nuestra segunda versión de la ley de los grandes números establece que el promedio $(1/n)\sum_{i=1}^{n} X_i$ converge en probabilidad a la media μ.

La ley Débil de los Grandes Números
Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes

e identicamente distribuidas (v.a.i.i.d.) tales que $E(X_i) = \mu$. Entonces para todo $\epsilon > 0$

$$\lim_{n \to \infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| \geq \epsilon \right) = 0.$$

Demostración. Aplicaremos la desigualdad de Chebyshev para hacer esta demostración. Supondremos adicionalmente que $\text{Var}(X_i) = \sigma^2 < \infty$. Sea

$$S_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Entonces $E(S_n) = \mu$ y $\text{Var}(S_n) = \sigma^2/n$. La desigualdad de Chebyshev aplicada a la variable aleatoria S_n dice que para toda $\epsilon > 0$ se tiene que

$$P\left(|S_n - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Basta ahora tomar el limite cuando n tiende a infinito para obtener el resultado requerido. •

5.4 El Teorema de De Moivre-Laplace

El siguiente resultado es una aproximación de la distribución normal a la distribución binomial. Es un caso particular del importante teorema del límite central que estudiaremos en la siguiente sección.

Teorema de De Moivre-Lapace
Supongamos que tenemos una serie de n ensayos independientes en cada uno de los cuales ocurre un cierto evento evento A con la misma probabilidad $p \in (0,1)$. Sea n_A el número de ocurrencias

del evento A en los n ensayos. Entonces para cualesquiera numeros reales $a \leq b$ tenemos que

$$\lim_{n\to\infty} P(a \leq \frac{n_A - np}{\sqrt{np(1-p)}} \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-\frac{1}{2}t^2)\, dt.$$

En otras palabras, el teorema anterior establece que para un gran número de ensayos n, la variable aleatoria

$$\frac{n_A - np}{\sqrt{np(1-p)}}$$

tiene una distribución normal estándar.

EJERCICIOS

1. La probabilidad de que un componente falle durante ciertas pruebas de confiabilidad es de $p = 0.05$. Use el teorema de De Moivre-Laplace para calcular una aproximación de la probabilidad de que al probar 100 componentes, el número de fallas sean
   (a) al menos 5.
   (b) menor a 5.
   (c) entre 5 y 10.
2. La probabilidad de ocurrencia de un evento en un ensayo es de 0.3. ¿ Cuál es la probabilidad de que la frecuencia relativa de este evento en 100 ensayos se encuentre entre 0.2 y 0.5 ?
3. Supongamos que la probabilidad ocurrencia de un cierto evento A en un ensayo es de 0.05. ¿ Cuántos ensayos independientes deberán realizarse para conseguir la ocurrencia del evento A en al menos en 5 ocasiones con probabilidad 0.9 ?

4. Se tiene un gran lote de artículos de los cuales el 10% es defectuoso. Un lote de estos artículos es rechazado si contiene al menos 10 artículos defectuosos. ¿ Cuántos artículos deben ser probados para que con probabilidad de 0.6 un lote sea rechazado ?

5. En un cierto experimento se tiene que la probabilidad de ocurrencia de un evento es de 0.4. ¿ Cuántos ensayos independientes se necesitan realizar de este experimento para que con probabilidad de 0.9 la frecuencia relativa del evento dado difiera de la probabilidad de ocurrencia de este evento en a lo mas 0.1 ?

6. La probabilidad de ocurrencia de un cierto evento en un ensayo es de 0.7. ¿ Cuál es la probabilidad de que este evento aparezca en la mayoría de una sucesión de 200 ensayos independientes?

7. Use el teorema de De Moivre-Laplace para demostrar que para un número n suficientemente grande de ensayos,

$$P(p-\epsilon \leq \frac{n_A}{n} \leq p+\epsilon) \approx \Phi\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{-\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right).$$

en donde $\frac{n_A}{n}$ es la frecuencia relativa del evento A cuya probabilidad de ocurrencia es $p \in (0,1)$, y $\epsilon > 0$.

5.5 El Teorema del Límite Central

Finalmente enunciaremos uno de los resultados más importantes en probabilidad.

Teorema del Límite Central
Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes

e identicamente distribuidas (v.a.i.i.d.) con $E(X_i) = \mu$ *y* $Var(X_i) = \sigma^2 < \infty$*. Sea*

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Entonces $E(S_n) = \mu$*,* $Var(S_n) = \sigma^2/n$ *y para todo* $x \in R$*,*

$$\lim_{n\to\infty} P\left(\frac{S_n - \mu}{\sqrt{\sigma^2/n}} \leq x\right) = P(Z \leq x) \quad \textit{en donde } Z \sim N(0,1).$$

Dicho de manera informal, el teorema del límite central dice que el promedio estandarizado de variables aleatorias tiene un compartimiento límite normal estándar.

Bibliografía

[1] Abramowitz M., Stegun I. () *Handbook of Mathematical Functions.*

[2] Blake I. F. (1979) *An Introduction to Applied Probability.* John Wiley & Sons.

[3] Blomm G., Holst L., Sandell D. (1994) *Problems and snapshots from the world of probability.* Springer-Verlag.

[4] Esparza S. N. (1987). *Elementos de Probabilidad.* Instituto Politécnico Nacional, México.

[5] Feller W. (1973). *Introducción a la Teoría de Probabilidades y sus Aplicaciones.* Editorial Limusa, México.

[6] Garza T. (1983). *Elementos de Cálculo de Probabilidades.* Universidad Nacional Autónoma de México, México.

[7] Gnedenko B. V. (1963). *The Theory of Probability.* MIR.

[8] Grimmett G. R. , Stirzaker D. R. (1982). *Probability and Random Processes.* Clarendon Press, Oxford.

[9] Harris B. (). *Theory of Probability.* Addison-Wesley Publishing Co.

[10] Hoel, Port, Stone (). *Introduction to Probability Theory.* Houghton Mifflin Company.

[11] Kolmogorov A. N. (1950) *Foundations of the Theory of Probability.* Chelsea Publishing Company, New York.

[12] Maistrov L. E. (1974). *Probability theory, a historical sketch.* Academic Press, New York.

[13] Mood A. M., Graybill F. A., Boes D. C. (1983). *Introduction to the Theory of Statistics - 3rd ed.* McGraw Hill.

[14] Miller I. , Miller M. (1999) *John E. Freund's Mathematical Statistics - 6th ed.* Prentice Hall, New Jersey.

[15] Ross S. (1994). *A First Course in Probability - 4th ed.* Macmillan Publishing Company, USA.

[16] Todhunter I. (1949). *A history of the mathematical theory of probability, from the time of Pascal to that of Laplace.* Chelsea Publishing Company, New York.