Curso elemental de
PROBABILIDAD Y ESTADÍSTICA

Luis Rincón
Departamento de Matemáticas
Facultad de Ciencias UNAM
Circuito Exterior de CU
04510 México DF

Versión: Diciembre 2007

Una versión actualizada del presente texto se encuentra disponible en formato electrónico en la dirección `http://www.matematicas.unam.mx/lars`

Prólogo

El presente texto constituye el material completo del curso semestral de Probabilidad y Estadística, impartido por el autor a alumnos de la licenciatura en ciencias de la computación en la Facultad de Ciencias de la UNAM. Contiene el temario básico para un curso elemental e introductorio a algunos temas tradicionales de la probabilidad y la estadística, así como una colección de ejercicios. Algunos de estos ejercicios aparecen a lo largo del texto como parte de la lectura, y una colección más extensa aparece al final del libro incluyendo algunas soluciones o sugerencias para resolverlos.

El texto está dirigido de manera general a alumnos de las distintas carreras de ingeniería, ciencias de la computación, y otras carreras científicas similares, cuyos programas de estudio contemplan un semestre introductorio a estos temas. Como es natural en este tipo de cursos, no se hace énfasis en el rigor matemático de la demostración de los resultados, sino en el uso, interpretación y aplicación de éstos. Como prerequisitos para una lectura provechosa de este material, se requiere, en determinados momentos, tener cierta familiaridad con algunos conceptos elementales de álgebra y del cálculo diferencial e integral. El texto fue escrito en el sistema LaTeX, y la mayoría de las ilustraciones fueron elaboradas usando el paquete pstricks.

El autor agradece cualquier comentario, sugerencia o corrección enviada al correo electrónico que aparece abajo.

Luis Rincón
Diciembre 2007
Ciudad Universitaria UNAM
lars@fciencias.unam.mx

Contenido

Parte 1

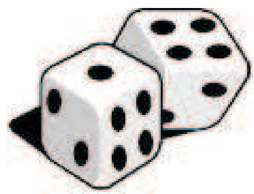
PROBABILIDAD

En esta primera mitad del curso estudiaremos algunos conceptos elementales de la teoría matemática de la probabilidad. Esta teoría tuvo como uno de sus primeros puntos de partida el intentar resolver un problema particular concerniente a una apuesta de juego de dados entre dos personas. El problema al que nos referimos involucraba una gran cantidad de dinero y puede plantearse de la siguiente forma:

> Dos jugadores escogen cada uno de ellos un número del 1 al 6, distinto uno del otro, y apuestan 32 doblones de oro a que el número escogido por uno de ellos aparece en tres ocasiones antes que el número del contrario al lanzar sucesivamente un dado. Suponga que el número de uno de los jugadores ha aparecido dos veces y el número del otro una sola vez. ¿Cómo debe dividirse el total de la apuesta si el juego se suspende?

Uno de los apostadores, Antonio de Gombaud, popularmente conocido como el caballero De Mere, deseando conocer la respuesta al problema plantea a Blaise Pascal (1623-1662) la situación. Pascal a su vez consulta con Pierre de Fermat (1601-1665) e inician un intercambio de cartas a propósito del problema. Esto sucede en el año de 1654. Con ello se inician algunos esfuerzos por dar solución a éste y otros problemas similares que se plantean. Con el paso del tiempo se sientan las bases y las experiencias necesarias para la búsqueda de una teoría matemática que sintetice los conceptos y los métodos de solución de los muchos problemas particulares resueltos a lo largo de varios años.

Blaise Pascal
(Francia, 1623–1662)



Pierre de Fermat
(Francia, 1601–1665)

En el segundo congreso internacional de matemáticas, celebrado en la ciudad de Paris en el año 1900, el matemático David Hilbert (1862-1943) plantea 23 problemas matemáticos de importancia. Uno de estos problemas es el de encontrar axiomas o postulados a partir de los cuales se pueda construir una teoría matemática de la probabilidad. Aproximadamente treinta años después, en 1933, el matemático ruso A. N. Kolmogorov (1903-1987) propone ciertos axiomas que a la postre resultaron adecuados para la construcción de una teoría de la probabilidad. Esta teoría prevalece hoy en día y ha adquirido el calificativo de teoría clásica. Actualmente la teoría clásica de la probabilidad se ha desarrollado y extendido enormemente gracias a muchos pensadores que han contribuído a su crecimiento, y es sin duda una parte importante y bien establecida de las matemáticas. Ha resultado útil para resolver problemas puramente matemáticos, pero sobre todo y principalmente, para modelar situaciones reales o imaginarias, en donde el azar es relevante.

1.1. Introducción

La teoría de la probabilidad es la parte de las matemáticas que se encarga del estudio de los fenómenos o experimentos aleatorios. Por *experimento aleatorio* entenderemos todo aquel experimento que cuando se le repite bajo las mismas condiciones iniciales, el resultado que se obtiene no siempre es el mismo. El ejemplo más sencillo y cotidiano de un experimento aleatorio es el de lanzar una moneda o un dado, y aunque estos experimentos pueden parecer muy modestos, hay situaciones en donde se utilizan para tomar decisiones de cierta importancia. En principio no sabemos cuál será el resultado del experimento aleatorio, asi que por lo menos conviene agrupar en un conjunto a todos los resultados posibles. El *espacio muestral* (o también llamado *espacio muestra* de un experimento aleatorio es el conjunto de todos los posibles resultados del experimento, y se le denota generalmente por la letra griega Ω (omega). Más adelante mostraremos que este conjunto no es necesariamente único y su determinación depende del interés del observador o persona que realiza el experimento aleatorio. En algunos textos se usa también la letra S para denotar al espacio muestral. Esta letra proviene del término *sampling space* de la lengua inglesa equivalente a *espacio muestral*. Por otro lado, llamaremos *evento* a cualquier subconjunto del espacio muestral y denotaremos a los eventos por las primeras letras del alfabeto en mayúsculas: A, B, C, etc. Con la ayuda de algunos ejemplos ilustraremos a continuación los conceptos de espacio muestral y evento.

Ejemplo. Si un experimento aleatorio consiste en lanzar un dado y observar el número que aparece en la cara superior, entonces claramente el espacio muestral es el conjunto $\Omega = \{1, 2, 3, 4, 5, 6\}$. Como ejemplo de un evento para este experimento podemos definir el conjunto $A = \{2, 4, 6\}$, que corresponde al suceso de obtener como resultado un número par. Si al lanzar el dado una vez se obtiene el número "4", decimos entonces que se observó la ocurrencia del evento A, y si se obtiene por ejemplo el resultado "1", decimos que no se observó la ocurrencia del evento A. ▪

Ejemplo. Considere el experimento aleatorio de participar en un juego de lotería. Suponga que hay un millón de números en esta lotería y un jugador participa con un boleto. ¿Cuál es un posible espacio muestral para este experimento? Naturalmente al jugador le interesa conocer su suerte en este juego y puede proponer como espacio muestral el conjunto $\Omega = \{$"ganar", "perder"$\}$. Sin embargo puede también tomarse como espacio muestral el conjunto que contiene a todos los posibles números ganadores, es decir, $\Omega = \{1, 2, \ldots, 1000000\}$. Este ejemplo sencillo muestra que el espacio muestral de un experimento aleatorio no es único y depende del interés del observador. ▪

Ejemplo. Suponga que un experimento aleatorio consiste en observar el tiempo en el que una máquina en operación sufre su primera descompostura. Si se consideran mediciones continuas del tiempo, entonces puede adoptarse como espacio muestral para este experimento el intervalo $[0, \infty)$. El subconjunto $A = [1, 2]$ corresponde al evento en el que la primera descompostura se observe entre la primera y la segunda unidad de tiempo. ▪

Ejercicio. Encuentre un espacio muestral para el experimento aleatorio de observar el marcador final de un juego de fútbol soccer. ¿Es un espacio muestral finito o infinito? ▪

Ejercicio. Suponga que se tiene en operación una sala de cómputo con 100 computadoras. Encuentre un espacio muestral para el experimento de observar la configuración de máquinas, desde el punto de vista de uso o no uso, en un momento cualquiera del día. ▪

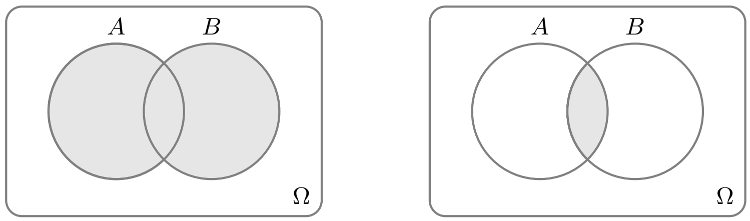
Puesto que los conceptos de espacio muestral y evento involucran forzosamente la terminología de conjuntos, recordaremos a continuación algunas operaciones entre estos objetos, y algunas propiedades que nos serán de utilidad en el estudio de la probabilidad y la estadística.

Operaciones con conjuntos. Supondremos que el espacio muestral Ω de un experimento aleatorio es una especie de conjunto universal, y cualquier elemento de Ω lo denotaremos por ω (omega minúscula). El conjunto vacío lo denotaremos por $\emptyset$. Otros símbolos usuales son los de pertenencia ($\in$), o no pertenencia ($\notin$) de un elemento en un conjunto, y los de contención ($\subset, \subseteq$), o no contención ($\not\subset$) de un conjunto en otro. Si A es un conjunto, denotamos la cardinalidad o número de elementos de ese conjunto por el símbolo $\#A$. Sean A y B dos subconjuntos cualesquiera de Ω. Recordamos a continuación las operaciones básicas de unión, intersección, diferencia y complemento:

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

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". En la Figura 1.1 se muestran en diagramas de Venn estas dos operaciones.

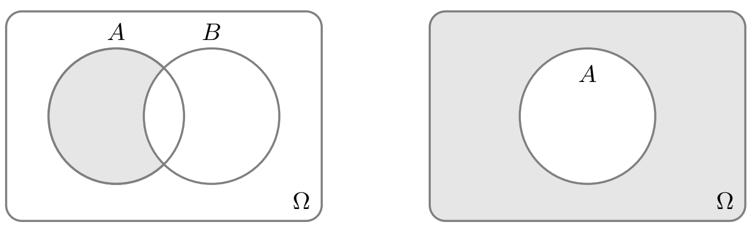
La diferencia entre dos conjuntos A y B se denota por $A - B$, y corresponde a aquel conjunto de elementos de A que no pertenecen a B, es decir, $A - B$ se define como $A \cap B^c$. En general, el conjunto $A - B$ es distinto de $B - A$, de hecho estos conjuntos son siempre ajenos. ¿Puede usted comprobar tal afirmación? ¿En qué caso ambos conjuntos coinciden? Por otro lado el complemento de un conjunto A se denota



$A \cup B$ $\quad$ $A \cap B$

Figura 1.1:

por A^c y se define como la colección de aquellos elementos de Ω que no pertenecen a A. Mediante un diagrama de Venn ilustramos gráficamente las operaciones de diferencia y complemento en la Figura 1.2.



$A - B$ $\quad$ A^c

Figura 1.2:

Ejemplo. Sea A el conjunto de aquellas personas que tienen hijos, y B la colección de aquellas personas que estan casadas. Entonces el conjunto $A \cap B$ consta de aquellas personas que estan casadas y tienen hijos, mientras que el conjunto $A \cap B^c$ está constituido por aquellas personas que tienen hijos pero no estan casadas. ¿Quién es $A^c \cap B$? Observe que cada persona es un elemento de alguno de los siguientes conjuntos: $A \cap B$, $A \cap B^c$, $A^c \cap B$ ó $A^c \cap B^c$. ¿A cuál pertenece usted? ∎

Ejercicio. Demuestre que si $A \subseteq B$, entonces $B^c \subseteq A^c$. ∎

Ejercicio. Considere los subconjuntos de números reales $A = \{x \in \mathbb{R} : x^2 \leq 1\}$ y $B = \{x \in \mathbb{R} : x > 0\}$. Encuentre y represente en el eje real los conjuntos: $A \cup B$,

$A \cap B$, $A - B$ y $B - A$. ∎

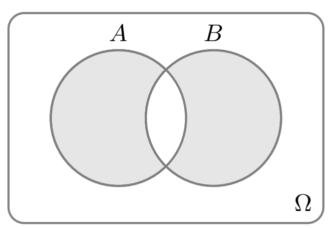
Es fácil verificar que el conjunto vacío y el conjunto total satisfacen las siguientes propiedades elementales: $A \cup \emptyset = A$, $A \cap \Omega = A$, $A \cap \emptyset = \emptyset$, $A \cup A^c = \Omega$, $A \cup \Omega = \Omega$, $A \cap A^c = \emptyset$. Las operaciones unión e intersección son *asociativas*, esto es, satisfacen las siguientes igualdades:

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

y también son *distributivas*, es decir,

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

Recordemos también la operación *diferencia simétrica* entre dos conjuntos A y B, denotada por $A \triangle B$, y definida como sigue: $A \triangle B = (A \cup B) - (B \cap A)$. En la Figura 1.3 ilustramos gráficamente el conjunto resultante de efectuar la diferencia simétrica entre los conjuntos A y B. Visualmente es fácil comprobar que la diferencia simétrica también puede escribirse como $(A - B) \cup (B - A)$. ¿Cómo podría expresarse en palabras al conjunto $A \triangle B$?



$A \triangle B$

Figura 1.3:

Recordemos además las muy útiles *leyes de De Morgan*:

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

La validez de estas dos igualdades puede extenderse a colecciones finitas e incluso arbitrarias de conjuntos. ¿Puede usted escribir estas identidades para n conjuntos?

Conjuntos ajenos. Decimos que dos conjuntos A y B son *ajenos* (o *disjuntos*) si se cumple la igualdad $A \cap B = \emptyset$, es decir, los conjuntos A y B son ajenos cuando no existe un elemento que pertenezca tanto a A como a B. Por ejemplo, si $\Omega = \{1, 2, 3, 4, 5, 6\}$, entonces los conjuntos $A = \{1, 2\}$ y $B = \{3, 4\}$ son ajenos pues no hay ningún elemento común entre ellos. El ejemplo general más importante de conjuntos o eventos ajenos es la pareja dada por A y A^c, para cualquier conjunto A.

Ejercicio. Suponga que A es el conjunto de raíces de la ecuación cuadrática $x^2 - x - 2 = 0$, y B es el intervalo $[0, 3]$. ¿Son los conjuntos A y B ajenos? ▪

Este concepto puede extenderse al caso cuando se tienen varios conjuntos. Decimos que n conjuntos $A_1, \ldots, A_n$ son *ajenos* si $A_1 \cap \cdots \cap A_n = \emptyset$, y se dice que son *ajenos dos a dos* (o *mutuamente ajenos*) si $A_i \cap A_j = \emptyset$ para cualesquiera valores de los índices $i, j = 1, 2, \ldots, n$, con i distinto de j. Existe diferencia en estas dos definiciones y explicaremos la situación con el siguiente ejemplo: Los conjuntos $A = \{1, 2\}$, $B = \{2, 3\}$ y $C = \{3, 4\}$ son ajenos pues $A \cap B \cap C = \emptyset$, pero no son ajenos dos a dos pues, por ejemplo, el conjunto $A \cap B$ no es vacío. Es decir, estos conjuntos son ajenos en el sentido de que la intersección de todos ellos es vacía pero no son ajenos dos a dos.

Ejercicio. Sean A y B dos conjuntos cualesquiera. Compruebe que la colección de conjuntos: $A \cap B$, $A^c \cap B$, $A \cap B^c$ y $A^c \cap B^c$ son ajenos dos a dos. Dibujar un diagrama de Venn podría ayudar a entender la situación. ▪

Ejercicio. Compruebe matemáticamente que si n conjuntos son ajenos, entonces son ajenos dos a dos. ▪

Las siguientes operaciones entre conjuntos no son elementales y producen nuevos conjuntos que se encuentran en un nivel distinto al de los conjuntos originales.

Conjunto potencia. El *conjunto potencia* de Ω, denotado por 2^Ω, es aquel conjunto cuyos elementos son todos los subconjuntos posibles de Ω. Por ejemplo, si $\Omega = \{a, b, c\}$, entonces el conjunto 2^Ω consta de 8 elementos, a saber,

$$2^\Omega = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}.$$

Observe que los elementos del conjunto potencia son en si mismos conjuntos, y que en esta colección estan contenidos todos los eventos que podrían ser de interés en un experimento aleatorio. No es difícil demostrar que $\#(2^{\Omega}) = 2^{\#\Omega}$, es decir, el número de elementos en el conjunto 2^{Ω} es exactamente 2 elevado a la potencia dada por la cardinalidad de Ω. De este hecho proviene la notación usada para el conjunto potencia: 2^{Ω}. Observe que la expresión 2^{Ω} no tiene sentido matemático, y debe considerarse como un símbolo para denotar al conjunto potencia. Para el ejemplo anterior se comprueba que la cardinalidad de 2^{Ω} es efectivamente $2^{\#\Omega} = 2^3 = 8$.

Ejercicio. Sea $\Omega = \{a\}$. ¿Quién es 2^{Ω}? ∎

Producto Cartesiano. Finalmente recordemos que el *producto Cartesiano* de dos conjuntos A y B, denotado por $A \times B$, se define como la colección de todas las parejas ordenadas (a, b), en donde a es cualquier elemento de A, y b es cualquier elemento de B. En símbolos, $A \times B = \{(a, b) : a \in A \text{ y } b \in B\}$. Por ejemplo, si $A = \{a_1, a_2\}$ y $B = \{b_1, b_2, b_3\}$, entonces

$$A \times B = \{(a_1, b_1), (a_1, b_2), (a_1, b_3), (a_2, b_1), (a_2, b_2), (a_2, b_3)\}.$$

En general los conjuntos producto $A \times B$ y $B \times A$ son distintos pues la pareja (a, b) es distinta de (b, a), sin embargo ambos conjuntos tienen la misma cardinalidad, esto es, ambos tienen el mismo número de elementos. Más aún, si la cardinalidad de A es el número n, y la cardinalidad de B es m, entonces la cardinalidad del conjunto $A \times B$ es el producto $n \cdot m$. Este resultado es llamado *principio de multiplicación* que usaremos más adelante. Un poco más generalmente, puede considerarse el producto Cartesiano de n conjuntos y comprobarse que

$$\#(A_1 \times \cdots \times A_n) = \#A_1 \cdots \#A_n.$$

Ejemplo. El producto Cartesiano $\mathbb{R} \times \mathbb{R}$ es el conjunto de todas las parejas de números reales (x, y). A este conjunto producto se le denota usualmente por $\mathbb{R}^2$. Análogamente se construyen los conjuntos $\mathbb{R}^3$, $\mathbb{R}^4$, ..., $\mathbb{R}^n$, ... ∎

Ejercicio. Sea $A = \{0, 1\}$. ¿Cuántos elementos tiene $\underbrace{A \times \cdots \times A}_{n}$? ∎

Concluimos aquí nuestra rápida y breve revisión de conjuntos. Recordemos que estamos interesados en calcular probabilidades de los diferentes eventos, es decir, de subconjuntos del espacio muestral que se obtienen al estudiar los diversos experimentos aleatorios. En la siguiente sección estudiaremos algunas formas de definir matemáticamente la probabilidad de un evento.

1.2. Probabilidad

La probabilidad de un evento A, es un número real en el intervalo $[0,1]$ que denotaremos por $P(A)$, y representa una medida de la *frecuencia* con la que se observa la ocurrencia del evento A cuando se efectúa el experimento aleatorio en cuestión. Existen al menos cuatro definiciones de probabilidad las cuales explicamos a continuación.

Probabilidad clásica. Sea A un subconjunto de un espacio muestral Ω de cardinalidad finita. Se define la *probabilidad clásica* del evento A como el cociente:

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

en donde el símbolo #A denota la cardinalidad o número de elementos del conjunto A. Claramente esta definición es sólo válida para espacios muestrales finitos, pues forzosamente necesitamos suponer que el número de elementos en Ω es finito. Además, el espacio Ω debe ser *equiprobable*, pues para calcular la probabilidad de un evento A, únicamente necesitamos contar cuántos elementos tiene A respecto del total Ω, sin importar exactamente qué elementos particulares sean. Por lo tanto, esta definición de probabilidad presupone que todos los elementos de Ω son *igualmente probables* o tienen el mismo peso. Este es el caso por ejemplo de un dado equilibrado. Para este experimento el espacio muestral es el conjunto $\Omega = \{1,2,3,4,5,6\}$, y si deseamos calcular la probabilidad (clásica) del evento A correspondiente a obtener un número par, es decir, la probabilidad de $A = \{2,4,6\}$, entonces

$$P(A) = \frac{\#\{2,4,6\}}{\#\{1,2,3,4,5,6\}} = \frac{3}{6} = \frac{1}{2}.$$

Probabilidad frecuentista. Suponga que se realizan n repeticiones de un cierto experimento aleatorio y sea A un evento cualquiera. Denotemos por $n(A)$ el número de ocurrencias del evento A, en las n realizaciones del experimento. Se define entonces la *probabilidad frecuentista* de A como indica el siguiente límite

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

En este caso, debemos hacer notar que no es humanamente posible llevar a cabo una infinidad de veces el experimento aleatorio, de modo que en la práctica no es posible encontrar mediante este mecanismo la probabilidad de un evento cualquiera. Esta limitación hace que esta definición de probabilidad no sea enteramente formal, pero tiene algunas ventajas. Veamos un ejemplo concreto. Consideremos nuevamente el experimento aleatorio de lanzar un dado equilibrado y registrar la ocurrencia del evento A definido como el conjunto $\{2, 4, 6\}$. Después de lanzar el dado 20 veces se obtuvieron los siguientes resultados:

No.	Resultado	$n(A)/n$
1	3	0/1
2	6	1/2
3	2	2/3
4	1	2/4
5	4	3/5
6	6	4/6
7	3	4/7
8	4	5/8
9	2	6/9
10	5	6/10

No.	Resultado	$n(A)/n$
11	2	7/11
12	5	7/12
13	1	7/13
14	6	8/14
15	3	8/15
16	1	8/16
17	5	8/17
18	5	8/18
19	2	9/19
20	6	10/20

En la gráfica de la Figura 1.4 se muestra el singular comportamiento de este cociente a lo largo del tiempo, al principio se pueden presentar algunas oscilaciones pero eventualmente el cociente se estabiliza en un cierto número. Realizando un mayor número de observaciones del experimento, no es difícil creer que el cociente $n(A)/n$ se estabiliza en $1/2$ cuando n es grande y el dado es equilibrado. Se invita al lector intrigado a efectuar un experimento similar y corroborar esta interesante *regularidad estadística* con éste o cualquier otro experimento aleatorio de su interés.

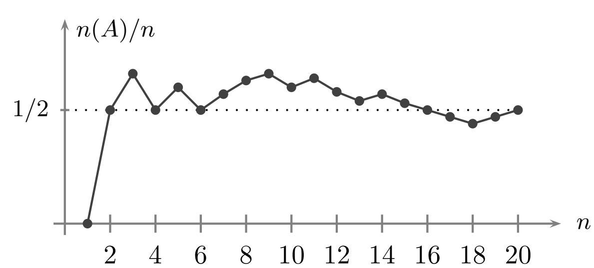


Figura 1.4:

Probabilidad subjetiva. En este caso la probabilidad de un evento depende del observador, es decir, según lo que el observador conoce del fenómeno en estudio. Puede parecer un tanto informal y poco seria esta definición de la probabilidad de un evento, sin embargo en muchas situaciones es necesario recurrir a un experto para tener por lo menos una idea vaga de cómo se comporta el fenómeno de nuestro interés y saber si la probabilidad de un evento es alta o baja. Por ejemplo, ¿Cuál es la probabilidad de que un cierto equipo de fútbol gane en su próximo partido? Ciertas circunstancias internas del equipo, las condiciones del equipo rival o cualquier otra condición externa, son elementos que sólo algunas personas conocen y que podrían darnos una idea más exacta de esta probabilidad. Esta forma subjetiva de asignar probabilidades a los distintos eventos debe, sin embargo, ser consistente con una serie de reglas naturales que estudiaremos a continuación.

Probabilidad axiomática. En la definición axiomática de la probabilidad no se establece la forma explícita de calcular las probabilidades sino únicamente se proponen las reglas que el cálculo de probabilidades debe satisfacer. Los siguientes tres postulados o axiomas[1] fueron establecidos en 1933 por el matemático ruso A. N. Kolmogorov.

A. N. Kolmogorov
(Rusia, 1903–1987)

Axiomas de la probabilidad

1. $P(A) \geq 0$.
2. $P(\Omega) = 1$.
3. $P(A \cup B) = P(A) + P(B)$,
   cuando $A \cap B = \emptyset$.

No es difícil verificar que las definiciones anteriores de probabilidad satisfacen estos tres axiomas. De hecho, estos postulados han sido tomados directamente del análisis

[1] Un *postulado* o *axioma* es una proposición que se acepta como válida y sobre la cual se funda una teoría.

cuidadoso y reflexivo de las definiciones de probabilidad mencionadas anteriormente. En particular, observe que el tercer axioma es válido no sólo para dos eventos ajenos sino para cualquier colección finita de eventos ajenos dos a dos. A cualquier función P que satisfaga los tres axiomas de Kolmogorov se le llama *medida de probabilidad*, o simplemente *probabilidad*. A partir de estos postulados es posible demostrar que la probabilidad cumple con una serie de propiedades interesantes.

Proposición. Para cualquier evento A, $\quad P(A^c) = 1 - P(A)$.

Demostración. De la teoría elemental de conjuntos tenemos que $\Omega = A \cup A^c$. Como A y A^c son eventos ajenos, por el tercer axioma, $P(\Omega) = P(A) + P(A^c)$. Finalmente, como $P(\Omega) = 1$, por el segundo axioma obtenemos $P(A^c) = 1 - P(A)$. □

Proposición. $P(\emptyset) = 0$.

Demostración. Como $\emptyset = \Omega^c$, usando la propiedad anterior, tenemos que $P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 0$. □

Las siguientes dos proposiciones suponen la situación $A \subseteq B$ que se muestra gráficamente en la Figura 1.5.

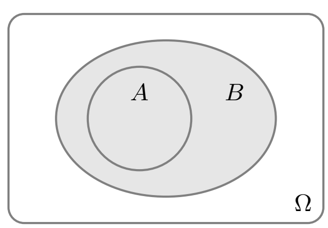


Figura 1.5: $A \subseteq B$.

Proposición. Si $A \subseteq B$, entonces $P(A) \leq P(B)$.

Demostración. Primeramente escribimos $B = A \cup (B - A)$. Como A y $B - A$ son eventos ajenos, por el tercer axioma, $P(B) = P(A) + P(B - A)$. Usando el primer axioma concluimos que $P(B) - P(A) = P(B - A) \geq 0$. De aqui obtenemos $P(B) - P(A) \geq 0$. □

Proposición. Si $A \subseteq B$, entonces $P(B-A) = P(B) - P(A)$.

Demostración. Como $B = A \cup (B-A)$, siendo esta unión ajena, por el tercer axioma tenemos que $P(B) = P(A) + P(B-A)$. □

Ejercicio. Sean $A \subseteq B$ eventos tales que $P(A^c) = 0.9$ y $P(B^c) = 0.6$. Compruebe que $P(B-A) = 0.3$. ▪

Proposición. Para cualquier evento A, $0 \leq P(A) \leq 1$.

Demostración. Como $A \subseteq \Omega$, se tiene que $P(A) \leq P(\Omega) = 1$. La otra desigualdad, $0 \leq P(A)$, es simplemente el primer axioma. □

Proposición. Para cualesquiera eventos A y B,

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

Demostración. Primeramente observamos que para cualesquiera eventos A y B se cumple la igualdad $A - B = A - (A \cap B)$. Entonces escribimos a $A \cup B$ como la unión disjunta de los siguientes tres eventos:

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

Ahora aplicamos la probabilidad. Por el tercer axioma,

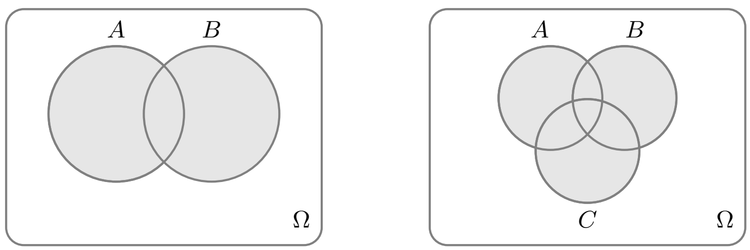
$$P(A \cup B) = P(A - A \cap B) + P(A \cap B) + P(B - A \cap B).$$

Pero $A \cap B \subseteq A$, de modo que $P(A - A \cap B) = P(A) - P(A \cap B)$. Análogamente $P(B - A \cap B) = P(B) - P(A \cap B)$. Por lo tanto

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

□

En la Figura 1.6 (a) el lector puede comprobar la validez de la fórmula anterior identificando las tres regiones ajenas de las que consta $A \cup B$. El término $P(A)$ abarca las primeras dos regiones de izquierda a derecha, $P(B)$ abarca la segunda y tercera región. Observe entonces que la región central ha sido contada dos veces de modo que el término $-P(A \cap B)$ da cuenta de ello. De esta forma las tres regiones son contadas una sola vez y el resultado es la probabilidad del evento $A \cup B$.



(a) $A \cup B$ (b) $A \cup B \cup C$

Figura 1.6:

Ejercicio. Sean A y B eventos ajenos tales que $P(B) = 0.3$ y $P(A \cap B^c) = 0.2$. Compruebe que $P(A \cup B) = 0.5$. ▪

Observe que la fórmula anterior es válida para cualesquiera eventos A y B. En particular, cuando son conjuntos ajenos, es decir, cuando $A \cap B = \emptyset$, entonces la fórmula demostrada se reduce al tercer axioma de la probabilidad, es decir, $P(A \cup B) = P(A) + P(B)$. El siguiente resultado es una generalización del anterior e involucra tres eventos cualesquiera. La fórmula que a continuación se demuestra puede también verificarse usando el diagrama de Venn que aparece en la Fig 1.6 (b). Para ello siga los términos del lado derecho de la fórmula y compruebe que cada región es contada una sola vez de modo que el resultado final es la probabilidad del evento $A \cup B \cup C$.

Proposición. Para cualesquiera eventos A, B y C,

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

Demostración. Usando la fórmula para dos eventos y agrupando adecuadamente,

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

□

Las propiedades anteriores son parte del estudio teórico y general de la probabilidad. En general, supondremos que la forma explícita de calcular estos números es conocida, o que se puede suponer un cierto modelo para llevar a cabo estos cálculos dependiendo del experimento aleatorio en cuestión. Por ejemplo, cuando el espacio muestral es finito y cada resultado puede suponerse igualmente probable, entonces usaremos la definición clásica de probabilidad. En otras situaciones asignaremos probabilidades de acuerdo a ciertos modelos conocidos. Regresaremos a este punto más adelante. A manera de resumen presentamos a continuación una tabla con las propiedades de la probabilidad que hemos demostrado.

Algunas propiedades de la probabilidad

a) $P(A^c) = 1 - P(A)$.

b) $P(\emptyset) = 0$.

c) Si $A \subseteq B$, entonces $P(A) \leq P(B)$.

d) Si $A \subseteq B$, entonces $P(B - A) = P(B) - P(A)$.

e) $0 \leq P(A) \leq 1$.

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

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

Esperamos que, a partir de las propiedades enunciadas y demostradas, el lector haya desarrollado cierta habilidad e intuición para escribir la demostración de alguna otra propiedad de la probabilidad. Otras propiedades sencillas pueden encontrarse en la sección de ejercicios. Debemos también decir que las demostraciones no son únicas, y que es altamente probable que el lector pueda producir alguna demostración diferente a las que aquí se han presentado.

1.3. Análisis combinatorio

Consideraremos ahora el caso cuando el experimento aleatorio es tal que su espacio muestral es un conjunto finito y cada elemento de este conjunto tiene la misma probabilidad de ocurrir, es decir, cuando el espacio Ω es finito y equiprobable. En estos casos hemos definido la probabilidad clásica de un evento A de la siguiente forma $P(A) = \#A/\#\Omega$. Para poder aplicar esta definición necesitamos saber contar cuántos elementos tiene un evento A cualquiera. Cuando podemos poner en una lista todos y cada uno de los elementos de dicho conjunto, entonces es fácil conocer la cardinalidad de A, simplemente contamos todos los elementos uno por uno. Sin embargo, es común enfrentar situaciones en donde no es factible escribir en una lista cada elemento de A. Por ejemplo, ¿Cuántos números telefónicos existen que contengan por lo menos un cinco? Es poco factible que alguien intente escribir uno a uno todos estos números telefónicos. En las siguientes secciones estudiaremos algunas técnicas de conteo que nos ayudarán a calcular la cardinalidad de un evento A en ciertos casos particulares. El *principio de multiplicación* que enunciamos a continuación es la base de muchos de los cálculos en las técnicas de conteo.

Proposición. Si un procedimiento A_1 puede efectuarse de n formas distintas y un segundo procedimiento A_2 puede realizarse de m formas diferentes, entonces el total de formas en que puede efectuarse el primer procedimiento seguido del segundo es el producto $n \cdot m$, es decir, $\#(A_1 \times A_2) = \#A_1 \cdot \#A_2$.

Para ilustrar este resultado considere el siguiente ejemplo. Suponga que un cierto experimento aleatorio consiste en seleccionar un dado y después seleccionar al azar una letra del alfabeto. ¿Cuál es la cardinalidad del correspondiente espacio muestral? El experimento de lanzar un dado tiene 6 resultados posibles y consideremos que tenemos un alfabeto de 26 letras. El correspondiente espacio muestral tiene entonces cardinalidad $6 \times 26 = 156$. El principio de multiplicación es válido no solamente para dos procedimientos sino que también vale para cualquier sucesión finita de procedimientos. Por ejemplo, si $A_1, A_2, \ldots, A_k$ denotan k procedimientos sucesivos, entonces este principio se puede enunciar en símbolos de la forma siguiente: $\#(A_1 \times \cdots \times A_k) = \#A_1 \cdots \#A_k$.

Ejercicio. Un hombre tiene 4 pantalones distintos, 6 camisas, y dos pares de zapatos. ¿De cuántas formas distintas puede el hombre vestirse con estas prendas? ▪

Vamos a considerar a continuación diferentes esquemas y contextos en donde es posible encontrar una fórmula matemática para ciertos problemas de conteo. En todos ellos aplicaremos el principio de multiplicación. El esquema general es el de extraer al azar k objetos, uno a la vez, de una urna con n objetos distintos. Esto se muestra en la Figura 1.7.

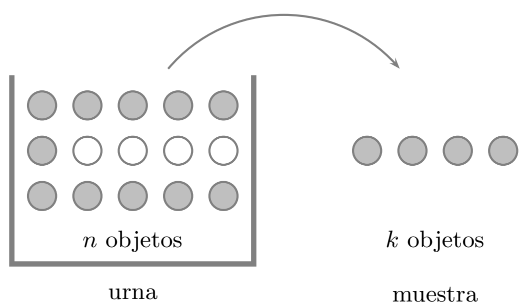


Figura 1.7:

ORDENACIONES CON REPETICIÓN: MUESTRAS CON ORDEN Y CON REEMPLAZO. Suponga entonces que tenemos una urna con n objetos distintos. Deseamos realizar k extracciones al azar de un objeto a la vez. Al efectuar una extracción, registramos el objeto escogido y lo regresamos a la urna, de esta forma el mismo objeto puede ser extraído varias veces. El total de arreglos que se pueden obtener de esta urna al hacer k extracciones es el número n^k, pues en cada extracción tenemos n objetos posibles para escoger y efectuamos k extracciones. Esta fórmula es consecuencia del principio de multiplicación enunciado antes. A este número se le llama *ordenaciones con repetición*. Se dice que la muestra es con *orden* pues es importante el orden en el que se van obteniendo los objetos, y es *con reemplazo* pues cada objeto seleccionado se reincorpora a la urna.

Ejemplo. Suponga que tenemos un conjunto de 60 caracteres diferentes que contiene todas las letras minúsculas del alfabeto, las letras mayúsculas, los diez dígitos y algunos caracteres especiales. ¿Cuántos passwords o palabras clave de longitud 4 se pueden construir usando el conjunto de 60 caracteres? Este es un ejemplo de una ordenación de 60 caracteres en donde se permiten las repeticiones. Como cada caracter de los 60 disponibles puede ser escogido para ser colocado en cada una de las cuatro posiciones de la palabra clave, entonces se pueden construir $60 \times 60 \times 60 \times 60 = 60^4 = 12,960,000$ distintos passwords de longitud 4. ▪

Ejercicio. Diez personas votarán por uno de tres candidatos. ¿De cuántas formas distintas pueden los votos ser registrados uno por uno? Solución: 59049. ▪

Ejercicio. En un partido de fútbol hubo 4 goles en el marcador final. ¿Cuántas historias distintas de marcadores llevan a un marcador final donde hay 4 goles? Solución: 16. ▪

Ordenaciones sin repetición: Muestras con orden y sin reemplazo. Suponga que se tiene la misma situación que antes, una urna con n objetos y de los cuales se deben extraer, uno a uno, k objetos. Suponga esta vez que el muestreo es *sin reemplazo*, es decir, una vez seleccionado un objeto éste ya no se reincorpora a la urna. El total de arreglos distintos que se pueden obtener de este modo es el número: $n(n-1)(n-2)\cdots(n-k+1)$. Primeramente debemos observar que hay k factores en la expresión anterior. El primer factor es n y ello es debido a que tenemos cualesquiera de los n objetos para ser colocado en primera posición, para la segunda posición tenemos ahora $n-1$ objetos, para la tercera $n-2$ objetos, etc. Este razonamiento termina al escoger el k-ésimo objeto para cual tenemos únicamente $n-k+1$ posibilidades. Nuevamente por el principio multiplicativo, la respuesta es el producto indicado. La expresión encontrada puede escribirse como sigue:

$$P(n,k) = \frac{n!}{(n-k)!},$$

y se lee *permutaciones de n en k*. En el caso particular cuando la muestra es exhaustiva, es decir, cuando $k = n$, o bien cuando todos los objetos son extraídos uno por uno, entonces se tienen todas las permutaciones o distintos órdenes en que se pueden colocar n objetos.

Ejemplo. ¿De cuantas formas distintas pueden asignarse los premios primero, segundo y tercero en una rifa de 10 boletos numerados del 1 al 10? Claramente se trata de una ordenación sin repetición de 10 objetos en donde se deben extraer 3 de ellos. La respuesta es entonces que existen $10 \times 9 \times 8 = 720$ distintas asignaciones para los tres primeros lugares en la rifa. ▪

Ejercicio. ¿De cuántas formas distintas pueden dos equipos de fútbol terminar en la clasificación general de un torneo en donde compiten 20 equipos? Solución: 380. ▪

Permutaciones: Muestras exhaustivas con orden y sin reemplazo. La pregunta básica acerca del total de formas en que podemos poner en orden lineal (uno detrás de otro y por lo tanto no hay repetición) n objetos distintos tiene como respuesta el *factorial* de n, denotado por $n!$ y definido como sigue:

$$n! = n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1.$$

A este número también se le conoce como las *permutaciones* de n objetos, y se usa la notación $P(n) = n!$ Adicionalmente y por conveniencia se define $0! = 1$. Observe que las permutaciones de n objetos es un caso particular de la situación mencionada en la sección anterior sobre ordenaciones sin repetición pero ahora cuando la muestra es exhaustiva, es decir, cuando se extraen los n objetos de la urna.

Ejemplo. Si deseamos conocer el total de formas distintas en que podemos colocar una enciclopedia de 5 volúmenes en un librero, la respuesta es claramente $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$. El razonamiento es el siguiente: Cualquiera de los cinco libros puede ser colocado al principio, quedan cuatro libros por colocar en la segunda posición, restan entonces tres posibilidades para la tercera posición, etc. Por el principio de multiplicación la respuesta es el producto de estos números. ▪

Ejercicio. La novela *Rayuela* del escritor argentino Julio Cortázar contiene 56 capítulos que aparentemente pueden ser leídos en cualquier orden. ¿De cuántas formas distintas puede ser leída esta novela? ▪

Combinaciones: Muestras sin orden y sin reemplazo. Supongamos nuevamente que tenemos un conjunto de n objetos distinguibles y nos interesa obtener una muestra de tamaño k. Supongamos ahora que las muestras deben ser *sin orden* y *sin reemplazo*. Es decir, en la muestra no debe haber elementos repetidos, pues no hay reemplazo, y además la muestra debe verse como un conjunto pues no debe haber orden entre sus elementos. ¿Cuántas diferentes muestras podemos obtener de estas características? Para responder a esta pregunta seguiremos el siguiente razonamiento. Cuando el orden importa hemos encontrado antes la fórmula

$$\frac{n!}{(n-k)!}.$$

Ahora que no nos interesa el orden, observamos que cada uno de los arreglos de la fórmula anterior, está siendo contado $k!$ veces, las veces en que los mismos k elementos pueden ser permutados unos con otros, siendo que el conjunto de elementos es el mismo. Para obtener arreglos en donde el orden no importa, debemos entonces dividir por $k!$ La fórmula a la que hemos llegado se llama *combinaciones de n en k*, que denotaremos como sigue:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

A este número también se le conoce con el nombre de *coeficiente binomial de n en k*, pues aparece en el famoso *teorema del binomio*:

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

Para los casos $n=2$ y $n=3$ el teorema del binomio se reduce a las siguientes fórmulas que muy seguramente el lector conoce:

$$\begin{aligned} (a+b)^2 &= a^2 + 2ab + b^2. \\ (a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3. \end{aligned}$$

Ejemplo. ¿Cuántos equipos distintos de tres personas pueden escogerse de un grupo de 5 personas? Observe que el orden de las tres personas escogidas no es importante de modo que la respuesta es $\binom{5}{3} = 5!/(3!\,(5-3)!) = 10$. ▪

Ejercicio. En un popular juego de lotería se pide adivinar los seis números que serán escogidos al azar dentro del conjunto $\{1, 2, \ldots, 49\}$. ¿Cuál es el total de arreglos con los cuales un jugador puede participar en este juego? Si se establece que un jugador obtiene un segundo lugar si acierta únicamente a cinco de los seis números seleccionados, ¿cuántos segundos lugares puede haber para una arreglo dado de seis números? ▪

Ejercicio. Un cierto partido de fútbol terminó con un marcador de 4-4. ¿De cuántas formas distintas pudieron los equipos alternarse en anotar para llegar a tal resultado? Solución: 70. ¿Cree usted que todos estos resultados son igualmente probables? ∎

El coeficiente binomial es también una forma de generar las entradas del así llamado *triángulo de Pascal*, que puede observarse en Figura 1.8.

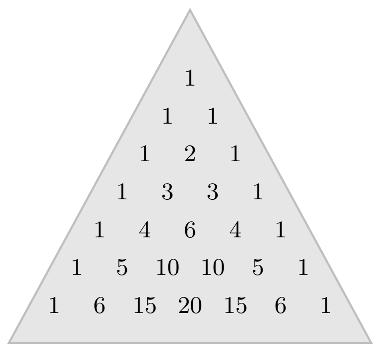


Figura 1.8:

El n-ésimo renglón del triángulo de Pascal, iniciando desde cero, contiene los coeficientes del desarrollo de $(a+b)^n$. Existe una forma sencilla de construir este triángulo observando que cada uno de estos números, exceptuando los extremos, es la suma de los dos números inmediatos del renglón anterior. A este respecto véase por ejemplo el Ejercicio 53 en la página 112.

Coeficiente multinomial. Ahora consideremos que tenemos n objetos no necesariamente distintos unos de otros. Por ejemplo, supongamos que tenemos k_1 objetos de un primer tipo, k_2 objetos de un segundo tipo, y así sucesivamente, hasta k_m objetos del tipo m, en donde $k_1 + k_2 + \cdots + k_m = n$. Estos n objetos pueden todos ordenarse uno detrás de otro de tantas formas distintas como indica el así llamado *coeficiente multinomial*:

$$\binom{n}{k_1\, k_2\, \cdots\, k_{m-1}\, k_m} = \frac{n!}{k_1!\, k_2! \cdots k_{m-1}!\, k_m!}.$$

Un razonamiento para obtener esta fórmula es el siguiente. Si consideramos que los n objetos son todos distintos, entonces claramente las distintas formas en que pueden escribirse todos estos objetos uno detrás de otro es $n!$ Pero para cada uno de estos arreglos, los k_1 objetos del primer tipo, supuestos inicialmente distintos cuando en realidad no lo son, pueden permutarse entre sí de $k_1!$ formas diferentes, siendo que el arreglo total es el mismo. De aqui que debamos dividir por $k_1!$ Lo mismo sucede con los elementos del segundo tipo y así sucesivamente hasta los elementos del tipo m. El coeficiente multinomial aparece en la siguiente fórmula:

$$(a_1 + a_2 + \cdots + a_m)^n = \sum \binom{n}{k_1 \cdots k_m} a_1^{k_1} a_2^{k_2} \cdots a_m^{k_m}, \tag{1.1}$$

en donde la suma se efectúa sobre todos los posibles valores enteros no negativos de $k_1, k_2, \ldots, k_m$, tales que $k_1 + k_2 + \cdots + k_m = n$. Por ejemplo, compruebe el lector que la fórmula (1.1) produce la siguiente expresión:

$$(a + b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc.$$

¿Puede usted desarrollar $(a + b + c)^3$? Es interesante observar que cuando hay únicamente dos tipos de objetos, el coeficiente multinomial se reduce al coeficiente binomial.

Muestras sin orden y con reemplazo. Finalmente consideremos el caso de hacer k extracciones de una urna de n objetos con las condiciones de que cada objeto extraído es regresado a la urna (y entonces puede ser elegido nuevamente), y en donde el orden de la muestra no es relevante. Para encontrar una fórmula para el total de muestras que pueden obtenerse con estas características usaremos una modelación distinta pero equivalente. Consideremos el arreglo de n casillas de la Figura 1.9 junto con la siguiente interpretación.

$\times\times$		$\times$	$\times$	$\cdots$		$\times$
1	2	3	4	$\cdots$	$n-1$	n

Figura 1.9:

La primera casilla tiene dos cruces y eso indica que la bola uno fue seleccionada dos veces, la segunda casilla esta vacía y ello significa que la bola dos no fue seleccionada, etc. El número de cruces en la casilla i indica entonces el número de veces que la bola i fue seleccionada. En total debe haber k cruces pues es el total de extracciones. Deseamos entonces conocer el número de posibles arreglos que pueden obtenerse con estas características, y debe ser claro, después de algunos momentos de reflexión, que éste es el número de muestras de tamaño k, con reemplazo y sin orden, que se pueden obtener de un conjunto de n elementos distinguibles. Consideremos que las dos paredes en los extremos de este arreglo son fijas, estas paredes se encuentran ligeramente remarcadas. Consideremos además que las posiciones intermedias, cruz o linea vertical, pueden moverse. En total hay $n+k-1$ objetos movibles y cambiar de posición estos objetos produce las distintas configuraciones posibles que nos interesan. El número total de estos arreglos es

$$\binom{n+k-1}{k}$$

que equivale a colocar dentro de las $n+k-1$ posiciones las k cruces, dejando en los lugares restantes las paredes movibles.

Resumen de fórmulas. En el contexto de muestras de tamaño k tomadas de un conjunto de cardinalidad n, y a manera de resumen parcial, tenemos la siguiente tabla de fórmulas. Debemos hacer énfasis sin embargo en que para resolver un problema de conteo en particular, no debemos clasificarlo forzosamente y de manera mecánica en alguno de los esquemas mencionados. Muy posiblemente el problema en cuestión requiera de un razonamiento especial que involucre alguna combinación de las fórmulas encontradas. A menudo los problemas de conteo son difíciles de resolver y en algunos casos uno puede encontrar dos o mas "soluciones" distintas y aparentemente correctas. A veces ello se debe a que el problema no esta especificado de manera completa.

Muestras	con reemplazo	sin reemplazo
con orden	n^k	$\frac{n!}{(n-k)!}$
sin orden	$\binom{n+k-1}{k}$	$\binom{n}{k}$

1.4. Probabilidad condicional e independencia

Los conceptos de probabilidad condicional e independencia surgieron de manera natural en el proceso de encontrar solución a algunos problemas provenientes de situaciones reales. En esta sección estudiaremos estos conceptos y demostraremos además dos resultados de amplia aplicación: el teorema de probabilidad total y el teorema de Bayes.

Probabilidad condicional. Sean A y B dos eventos en donde B es tal que su probabilidad es estrictamente positiva. La *probabilidad condicional* del evento A dado el evento B, denotada por $P(A|B)$, se define como sigue:

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

La expresión $P(A|B)$ se lee *probabilidad condicional del evento A dado el evento B*, o simplemente *probabilidad de A dado B*. Es claro que para que la definición tenga sentido se necesita suponer que $P(B) > 0$, y por otro lado no existe definición para $P(A|B)$ cuando $P(B) = 0$. Ilustraremos con un ejemplo el significado de la probabilidad condicional y comprobaremos que el evento B representa información adicional acerca del experimento aleatorio que modifica, en general, las probabilidades de los distintos eventos.

Ejemplo. Considere el experimento de lanzar un dado equilibrado. Claramente el espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$, el cual por hipótesis es equiprobable. Sean los eventos $A = \{2\}$ y $B = \{2, 4, 6\}$ = "Cae par". Entonces $P(A) = 1/6$ mientras que

$$P(A|B) = \frac{P(\{2\} \cap \{2, 4, 6\})}{P(\{2, 4, 6\})} = \frac{P(\{2\})}{P(\{2, 4, 6\})} = \frac{1/6}{3/6} = 1/3.$$

Observe que conocer la información de la ocurrencia del evento B, ha afectado la probabilidad del evento A, es decir, dada la información que el resultado del dado es un número par, la probabilidad de obtener "2" es ahora 1/3. ▪

Ejercicio. La siguiente fórmula se llama *regla del producto*: Sean $A_1, \ldots, A_n$ eventos tales que $P(A_1 \cap \cdots \cap A_{n-1}) > 0$. Compruebe que

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

▪

Ejercicio. El siguiente experimento se conoce como la *urna de Polya*. Suponga que en una urna se tienen b bolas blancas y r bolas rojas. Un experimento aleatorio consiste en seleccionar una bola al azar y regresarla a la urna junto con c bolas del mismo color. Use la regla del producto para calcular la probabilidad de obtener bolas rojas en las tres primeras extracciones. ▪

Independencia de eventos. Se dice que dos eventos cualesquiera A y B son *independientes* si se cumple la condición $P(A \cap B) = P(A)P(B)$. Esta igualdad es equivalente a la expresión $P(A|B) = P(A)$ cuando $P(B) > 0$. La ventaja de esta última expresión es que posee una interpretación sencilla: Dice que la probabilidad del evento A es la misma cuando sabemos que ha ocurrido el evento B (lado izquierdo), que cuando no sabemos nada (lado derecho). Es decir, la ocurrencia del evento B no afecta la probabilidad del evento A y por lo tanto son independientes. De manera análoga puede interpretarse la igualdad equivalente $P(B|A) = P(B)$, suponiendo naturalmente que $P(A) > 0$. En la mayoría de los casos de aplicación simplemente supondremos que dos eventos dados son independientes recurriendo únicamente a justificaciones intuitivas.

Ejemplo. Suponga que se lanza una moneda dos veces. Es una hipótesis natural suponer que el resultado del primer lanzamiento no afecta el resultado del segundo lanzamiento. De este modo cualquier evento del primer ensayo es independiente de cualquier otro evento en el segundo ensayo. ▪

La definición de independencia de dos eventos puede generalizarse al caso de varios eventos de la siguiente forma: Decimos que n eventos $A_1, A_2, \ldots, A_n$ son *independientes* si se satisfacen todas y cada una de las condiciones siguientes:

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

En general, para verificar que n eventos son independientes es necesario comprobar todas y cada una de las igualdades arriba enunciadas. Es decir, cualquiera de estas igualdades no implica, en general, la validez de alguna otra, es necesario pues verificarlas todas. No es difícil darse cuenta que el total de igualdades es $2^n - n - 1$. ¿Puede usted justificar este resultado?

Ejemplo. — ▪

Ejercicio. Sea $\Omega = \{1, 2, 3, 4\}$ un espacio muestral equiprobable. Sean los eventos $A = \{1, 2\}$, $B = \{2, 3\}$ y $C = \{2, 4\}$. ¿Son A, B y C independientes? ▪

Teorema de probabilidad total. Antes de enunciar el siguiente resultado recordaremos el concepto de partición de un conjunto. Una *partición* finita de un conjunto Ω es una colección $B_1, \ldots, B_n$ de subconjuntos de Ω tal que cada uno de estos conjuntos es distinto del vacío, la colección es disjunta dos a dos, esto es, para índices i y j distintos, se cumple que $B_i \cap B_j \neq 0$, y además la unión de toda la colección produce el total Ω, es decir, $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$. En la Figura 1.10 se muestra gráficamente el concepto de partición de un conjunto.

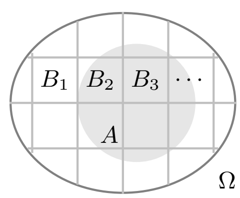


Figura 1.10:

Ahora podemos enunciar y demostrar el muy útil teorema de probabilidad total.

Teorema de probabilidad total. Sea $B_1, B_2, \ldots, B_n$ una partición de Ω tal que cada elemento de la partición tiene probabilidad estrictamente positiva. Para cualquier evento A,

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

Demostración. Primero observemos que el evento A puede escribirse como sigue

$$A = A \cap \Omega = A \cap \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A \cap B_i,$$

en donde los eventos $A \cap B_1, \ldots, A \cap B_n$ son ajenos dos a dos, de modo que

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

□

Observe que cuando la partición de Ω consta de únicamente dos elementos, B y B^c, la fórmula del teorema de probabilidad total se reduce a la siguiente expresión.

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

Ejemplo. Suponga que tenemos dos cajas: una con 3 bolas blancas y 7 de color gris, la otra con 6 blancas y 6 grises. Si se elije una caja al azar y después se saca una bola, ¿cuál es la probabilidad de que sea blanca?

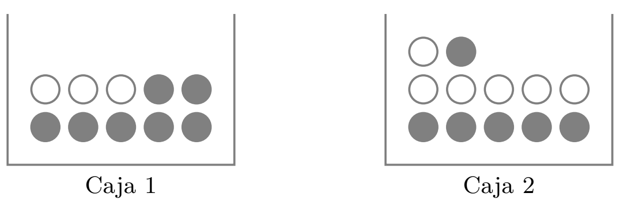


Figura 1.11:

El experimento aleatorio consiste entonces en escoger una caja al azar, con idéntica probabilidad cada una de ellas, y después escoger una bola de la caja escogida. Es claro entonces que el espacio muestral puede escribirse como sigue

$$\Omega = \{(C_1, B), (C_1, G), (C_2, B), (C_2, G)\},$$

en donde C_1 y C_2 denotan los eventos en donde las cajas uno y dos fueron escogidas, respectivamente, y B y G denotan los eventos en donde una bola blanca o gris fueron escogidas respectivamente. Nos piden calcular la probabilidad de B. Observe que es fácil calcular la probabilidad de este evento cuando se conoce la caja que fue escogida. Esto sugiere condicionar sobre el resultado de escoger alguna de las dos cajas y aplicar el teorema de probabilidad total, es decir,

$$\begin{aligned} P(B) &= P(B|C_1)P(C_1) + P(B|C_2)P(C_2) \\ &= \frac{3}{10} \times \frac{1}{2} + \frac{6}{12} \times \frac{1}{2} \\ &= \frac{2}{5}. \end{aligned}$$

Observe además que la partición del espacio muestral consta de dos elementos: $\{(C_1, B), (C_1, G)\}$ y $\{(C_2, B), (C_2, G)\}$. ¿Puede usted comprobar que $P(G) = 3/5$? Puede uno también preguntarse por situaciones aparentemente extrañas como la siguiente: Si se obtuvo una bola blanca, ¿Cuál es la probabilidad de que haya sido obtenida de la primera caja? ▪

Ejemplo. Suponga que en una población humana de igual número de hombres y mujeres, el 4 % de hombres son daltónicos y el 1 % de las mujeres son daltónicas. Una persona es elegida al azar, ¿Cuál es la probabilidad de que sea daltónica? Definamos primero los eventos de interés. Sea M el evento "La persona escogida es mujer", H el evento "La persona escogida es hombre" y D el evento "La persona escogida es daltónica". Deseamos calcular $P(D)$. Por el teorema de probabilidad total,

$$\begin{aligned} P(D) &= P(D|M)P(M) + P(D|H)P(H) \\ &= \frac{1}{100} \times \frac{1}{2} + \frac{4}{100} \times \frac{1}{2} \\ &= \frac{1}{40}. \end{aligned}$$

▪

Teorema de Bayes. Este es otro resultado interesante que involucra probabilidades condicionales. Fue publicado por primera vez en 1763, dos años después de la muerte de su creador, el matemático y teólogo inglés Thomas Bayes.

Teorema de Bayes. Sea $B_1, \ldots, B_n$ una partición de Ω tal que cada elemento de la partición tiene probabilidad estrictamente positiva. Sea A un evento tal que $P(A) > 0$. Entonces para cada $j = 1, 2, \ldots, n$,

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

Demostración. Por la definición de probabilidad condicional y el teorema de probabilidad total tenemos que

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

□

Véase la Figura 1.10 para una representación gráfica de la partición del espacio muestral y el evento A. Nuevamente observamos que en el caso cuando la partición de Ω consta de sólo dos elementos: B y B^c, el teorema de Bayes, para el evento B, adquiere la forma:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}.$$

Ejemplo. En una fábrica hay dos máquinas. La máquina 1 realiza el 60 % de la producción total y la máquina 2 el 40 %. De su producción total, la máquina 1 produce 3 % de material defectuoso, la 2 el 5 %. El asunto es que se ha encontrado un material defectuoso, ¿Cuál es la probabilidad de que este material defectuoso provenga de la máquina $\mathbb{M}_2$? Sea M_1 el evento "La máquina 1 produjo el material escogido", M_2 en evento "La máquina 2 produjo el material escogido" y finalmente sea D el evento "El material escogido es defectuoso". El problema es encontrar $P(M_2|D)$ y obervamos que la información que tenemos es $P(D|M_2)$. Por el teorema de Bayes tenemos entonces que

$$\begin{aligned}
P(M_2|D) &= \frac{P(D|M_2)P(M_2)}{P(D|M_1)P(M_1)+P(D|M_2)P(M_2)} \\
&= \frac{\frac{5}{100}\times\frac{40}{100}}{\frac{3}{100}\times\frac{60}{100}+\frac{5}{100}\times\frac{40}{100}} \\
&= \frac{10}{19}.
\end{aligned}$$

¿Puede usted calcular $P(M_1|D)$? ∎

Ejemplo. En un laboratorio se descubrió una prueba para detectar cierta enfermedad, y sobre la eficacia de dicha prueba se conoce lo siguiente: Si se denota por E el evento de que un paciente tenga la enfermedad y por N el evento de que la prueba resulte negativa, entonces se sabe que $P(N^c|E) = 0.95$, $P(N|E^c) = 0.96$ y $P(E) = 0.01$. Con esta información uno podría pensar que la prueba es muy buena, sin embargo calcularemos las probabilidades $P(E|N)$ y $P(E|N^c)$, usando el teorema de Bayes.

$$\begin{aligned}
P(E|N) &= \frac{P(N|E)P(E)}{P(N|E)P(E)+P(N|E^c)P(E^c)} \\
&= \frac{0.05\times 0.01}{0.05\times 0.01+0.96\times 0.99} \\
&= 0.000526\,.
\end{aligned}$$

Es bueno que esta probabilidad sea pequeña, pero por otro lado,

$$\begin{aligned}
P(E|N^c) &= \frac{P(N^c|E)P(E)}{P(N^c|E)P(E)+P(N^c|E^c)P(E^c)} \\
&= \frac{0.95\times 0.01}{0.95\times 0.01+0.04\times 0.99} \\
&= 0.193\,.
\end{aligned}$$

Esta última probabilidad es demasiado pequeña y por lo tanto la prueba no es muy confiable en tales casos. ▪

Ejercicio. El problema de las tres puertas (Monty Hall). Se le presentan a un concursante tres puertas cerradas detrás de una de las cuales hay un premio. El concursante debe adivinar la puerta que contiene el premio para ganarlo. Una vez que el concursante elige una puerta, y antes de abrirla, el presentador del concurso abre alguna de las puertas restantes y de la cual sabe que no contiene ningún premio. Entonces le pregunta al concursante si desea cambiar su decisión. ¿Qué debe hacer el concursante? Justifique su respuesta. ▪

1.5. Variables aleatorias

Dado un experimento aleatorio cualquiera, una *variable aleatoria* es una transformación X del espacio de resultados Ω al conjunto de números reales, esto es,

$$X : \Omega \to \mathbb{R}.$$

A menudo se escribe simplemente *v.a.* en lugar del término variable aleatoria. En sentido estricto una variable aleatoria es una función de Ω en $\mathbb{R}$ que satisface además cierta condición de medibilidad, pero omitiremos tales tecnicismos pues no son de utilidad para los propósitos de este curso. Suponga entonces que se efectúa el experimento aleatorio una vez y se obtiene un resultado ω en Ω. Al transformar este resultado con la variable aleatoria X se obtiene un número real $X(\omega) = x$. Podemos entonces suponer que los posibles resultados del experimento aleatorio son los diferentes números reales x que la función X puede tomar. Ilustramos de manera gráfica el concepto de variable aleatoria en la Figura1.12. Debemos hacer aqui la siguiente observación importante. Seguiremos la notación usual de usar la letra mayúscula X para denotar una variable aleatoria cualquiera, es decir, X es una función de Ω en $\mathbb{R}$, mientras que la letra minúscula x denota un número real y que es un posible valor de la variable aleatoria. En general, las variables aleatorias se denotan usando las últimas letras del alfabeto en mayúsculas, U, V, W, X, Y, Z, y para un valor cualquiera de ellas se usa la misma letra pero en minúscula.

Ejemplo. Suponga que un experimento aleatorio consiste en lanzar al aire una moneda y observar la cara superior una vez que la moneda cae. Denotemos por “Cara”

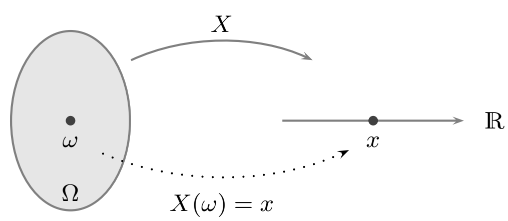


Figura 1.12:

y "Cruz" los dos lados de la moneda. Entonces claramente el espacio muestral es el conjunto $\Omega = \{$"Cara", "Cruz"$\}$. Defina la variable aleatoria $X : \Omega \to \mathbb{R}$ de la siguiente forma $X($"Cara"$) = 0$ y $X($"Cruz"$) = 1$. De este modo podemos suponer que el experimento aleatorio tiene dos valores numéricos posibles: 0 y 1. Observe que los números 0 y 1 son en realidad arbitrarios, otro par de números distintos puede ser escogido para distinguir los dos resultados del experimento aleatorio. Podemos también definir la variable aleatoria $Y : \Omega \to \mathbb{R}$ de la siguiente forma $Y($"Cara"$) = Y($"Cruz"$) = 2$. En este caso la variable Y solo toma un valor, el número 2. Cualquier resultado del experimento aleatorio produce, a través de la función Y, el número 2. Decimos que Y es la variable aleatoria constante 2. ▪

Ejemplo. Consideremos el experimento aleatorio consistente en lanzar un dardo en un tablero circular de radio uno. El espacio muestral o conjunto de posibles resultados del experimento se puede escribir como sigue $\Omega = \{(x, y) : x^2 + y^2 \leq 1\}$.

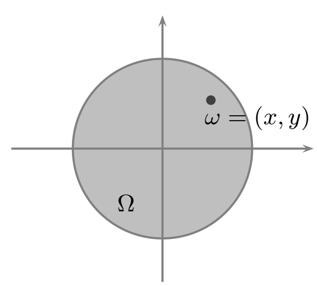


Figura 1.13:

Los siguientes son ejemplos de funciones de Ω en $\mathbb{R}$, variables aleatorias, asociadas a este experimento aleatorio.

a) $X(x,y) = x$, proyección sobre el eje horizontal.

b) $Y(x,y) = y$, proyección sobre el eje vertical.

c) $Z(x,y) = \sqrt{x^2+y^2}$, distancia al centro del círculo.

d) $V(x,y) = |x|+|y|$, distancia del taxista.

e) $W(x,y) = xy$, producto de las coordenadas. ▪

Considerando el conjunto de valores que una variable aleatoria puede tomar, vamos a clasificar a las variables aleatorias en dos tipos: discretas o continuas. Decimos que una v.a. es *discreta* cuando el conjunto de valores que ésta toma es un conjunto discreto, es decir, un conjunto finito o numerable. Por ejemplo, el conjunto $\{0,1,2,\ldots,n\}$ es un conjunto discreto porque es finito, lo mismo $\mathbb{N}$ pues aunque es infinito, es numerable y por lo tanto discreto. Por otra parte, decimos que una variable aleatoria es *continua* cuando toma todos los valores dentro de un intervalo $(a,b) \subseteq \mathbb{R}$. Esta clasificación de variables aleatorias no es completa pues existen variables que no son de ninguno de los dos tipos mencionados. Por simplicidad en este curso estudiaremos únicamente variables aleatorias que son discretas o continuas.

Ejemplo. Las variables X y Y definidas líneas arriba en el ejemplo del lanzamiento de una moneda, son variables aleatorias discretas. En ese ejemplo el espacio muestral mismo es discreto y por lo tanto las variables aleatorias que pueden allí definirse tienen que ser discretas forzosamente. En el ejemplo del lanzamiento de un dardo en un tablero circular de radio uno, el espacio muestral (Figura 1.13) es infinito no numerable, las variables X,Y,Z,V y W definidas allí son todas variables aleatorias continuas. Si se dibujan círculos concéntricos alrededor del origen y si se asignan premios asociados a cada una de las regiones resultantes, puede obtenerse un ejemplo de una variable aleatoria discreta sobre este espacio muestral. ▪

Ejemplo. Un experimento aleatorio consiste en escoger a una persona ω al azar. La variable aleatoria X evaluada en ω corresponde a conocer la siguiente característica, o una codificación de esta característica, de la persona escogida. En cada caso se trata de una variable aleatoria discreta: a) Edad en años. b) Número de hijos. c) Peso. d) Estatura. e) Sueldo. f) Nivel escolar. g) Estado civil. h) Lugar de nacimiento. ▪

Usaremos también la siguiente notación importante: Si A es un subconjunto de $\mathbb{R}$, entonces la expresión $(X \in A)$, incluyendo el paréntesis, denota el conjunto $\{\omega \in \Omega : X(\omega) \in A\}$, es decir, $(X \in A) = \{\omega \in \Omega : X(\omega) \in A\}$. En palabras, la expresión $(X \in A)$ denota aquel conjunto de elementos ω de Ω tales que bajo la aplicación de la función X toman un valor dentro del conjunto A. A este conjunto se le llama la *imagen inversa* de A, y se le denota por $X^{-1}A$.

Ejemplo. Consideremos nuevamente el experimento de lanzar una moneda y la variable aleatoria X que lleva el resultado "Cara" al valor 0 y el resultado "Cruz" al valor 1. Tenemos por ejemplo que $(X \in [1,\infty)) = \{\text{"Cruz"}\}$ pues el conjunto de elementos de Ω tales que bajo la función X toman un valor mayor o igual a uno, es decir caen dentro del intervalo $[1,\infty)$, es únicamente el elemento "Cruz". Por lo tanto $P(X \in [1,\infty)) = P\{\text{"Cruz"}\} = 1/2$. Del mismo modo puede verificarse que

a) $P(X \in [1,2)) = P(\{\text{"Cruz"}\}) = 1/2$.

b) $P(X \in [0,1)) = P(\{\text{"Cara"}\}) = 1/2$.

c) $P(X \in [2,4]) = P(\emptyset) = 0$.

d) $P(X = 1) = P(\{\text{"Cruz"}\}) = 1/2$.

e) $P(X \leq -1) = P(\emptyset) = 0$.

f) $P(X \geq 0) = P(\Omega) = 1$. ▪

Usaremos con mucha frecuencia la notación arriba explicada. El lector debe asegurarse de comprender bien que si x es un número real entonces $(X \leq x)$ es un subconjunto de Ω y por lo tanto un evento. Lo mismo sucede con el complemento de este conjunto que es $(X > x)$. Podemos escribir entonces la igualdad de conjuntos $(X \leq x) \cup (X > x) = \Omega$. Y aplicando probabilidad se obtiene $P(X \leq x) + P(X > x) = 1$.

Nota importante. A través de una variable aleatoria se puede considerar que los posibles resultados de un experimento aleatorio no son elementos ω en Ω sino números reales que la variable aleatoria puede tomar. Esta es una consideración radical pues ya no consideraremos experimentos aleatorios particulares, ni espacios muestrales arbitrarios Ω, ni eventos (subconjuntos) de Ω, en lugar de ello consideraremos que una cierta variable aleatoria de interés toma valores en un cierto subconjunto de números reales. La probabilidad definida antes para subconjuntos de Ω se traslada, como explicamos antes, a probabilidades para subconjuntos de $\mathbb{R}$. Esta perspectiva permite estudiar modelos generales y después aplicarlos a cualquier situación particular. A partir de ahora y en lo que resta del curso el término variable aleatoria constituirá un elemento frecuente en los enunciados.

1.6. Funciones de densidad y de distribución

En esta sección vamos a explicar la forma de asociar a cada variable aleatoria dos funciones que nos proveen de información acerca de las características de la variable aleatoria. Estas funciones, llamadas *función de densidad* y *función de distribución*, nos permiten representar a un mismo tiempo tanto los valores que puede tomar la variable como las probabilidades de los distintos eventos. Definiremos primero la función de densidad para una variable aleatoria discreta, después para una continua, y finalmente definiremos la función de distribución para ambos tipos de variables aleatorias.

Función de probabilidad para una variable discreta. Sea X una variable aleatoria discreta que toma los valores $x_1, x_2, \ldots$ con probabilidades respectivas $P(X = x_1), P(X = x_2), \ldots$. Esta lista de valores numéricos y sus probabilidades puede ser finita o bien infinita, pero numerable. La *función de probabilidad* de la variable X denotada por $f(x) : \mathbb{R} \to [0, \infty)$ se define como sigue

$$f(x) = \begin{cases} P(X = x) & \text{si } x = x_1, x_2, \ldots \\ 0 & \text{otro caso.} \end{cases} \tag{1.2}$$

En palabras, la función de probabilidad es simplemente aquella función que indica los valores de la probabilidad en los distintos valores que toma la variable aleatoria discreta. Recordemos que es importante poder distinguir entre X y x, pues conceptualmente son cosas muy distintas. Denotaremos generalmente a una función de probabilidad con la letra f minúscula. A veces escribiremos $f_X(x)$ y el subíndice nos ayudará a especificar que tal función es la función de probabilidad de la variable X. Esta notación será particularmente útil cuando consideremos varias variables aleatorias a la vez. Observe que se cumplen las siguientes dos propiedades, y toda función de la forma (1.2) la llamaremos *función de probabilidad*.

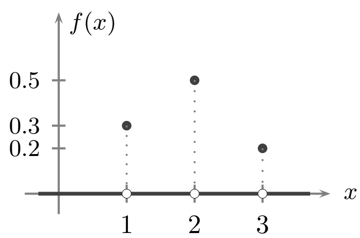
a) $f(x) \geq 0, \quad$ para toda $x \in \mathbb{R}$.

b) $\displaystyle\sum_x f(x) = 1$.

Ejemplo. Considere la variable aleatoria discreta X que toma los valores $1, 2$ y 3, con probabilidades 0.3, 0.5 y 0.2 respectivamente. Entonces la función de probabilidad de X es

$$f(x) = \begin{cases} 0.3 & \text{si } x = 1, \\ 0.5 & \text{si } x = 2, \\ 0.2 & \text{si } x = 3, \\ 0 & \text{otro caso.} \end{cases}$$

Esta función se muestra gráficamente en la Figura 1.14 (a). Alternativamente podemos también expresar esta función mediante la tabla de la Figura 1.14 (b). En esta representación se entiende de manera implícita que $f(x)$ es cero para cualquier valor de x distinto de 1, 2 y 3. En particular, compruebe que las siguientes probabilidades son correctas: $P(X \geq 2) = 0.7$, $P(|X| = 1) = 0.3$, y $P(X < 1) = 0$. ▪



(a)

x	1	2	3
$f(x)$	0.3	0.5	0.2

(b)

Figura 1.14:

Ejemplo. Encontraremos el valor de la constante c que hace que la siguiente función sea de probabilidad.

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

Los posibles valores de la variable aleatoria discreta, no especificada, son $0, 1, 2$ y 3, con probabilidades $0, c, 2c$ y $3c$, respectivamente. Como la suma de estas probabilidades debe ser uno, obtenemos la ecuación $c + 2c + 3c = 1$. De aqui obtenemos $c = 1/6$. Este es el valor de c que hace que $f(x)$ sea no negativa y sume uno, es decir, una función de probabilidad. ▪

Ejercicio. Grafique y compruebe que las siguientes funciones son de probabilidad.

a) $f(x) = x^2/10$, para $x = -2, -1, 0, 1, 2$.

b) $f(x) = (2x-5)^2/70, \quad$ para $\; x = 1,2,3,4,5.$

■

Función de densidad para una variable continua. Sea X una variable aleatoria continua. Decimos que la función integrable y no negativa $f(x) : \mathbb{R} \to [0,\infty)$ es la *función de densidad* de X si para cualquier intervalo (a,b) de $\mathbb{R}$ se cumple la igualdad

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

Es decir, la probabilidad de que la variable tome un valor dentro del intervalo (a,b) se puede calcular o expresar como el área bajo la función de densidad en el intervalo (a,b). De esta forma el cálculo de una probabilidad se reduce al cálculo de una integral. Véase la Figura 1.15. No es difícil comprobar que toda función de densidad $f(x)$ de una variable aleatoria continua cumple las siguientes propiedades análogas al caso discreto.

a) $f(x) \geq 0, \quad$ para toda $x \in \mathbb{R}$.

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

Toda función $f(x) : \mathbb{R} \to [0,\infty)$ que satisfaga estas dos propiedades, sin necesidad de tener una variable aleatoria de por medio, se llamará *función de densidad*.

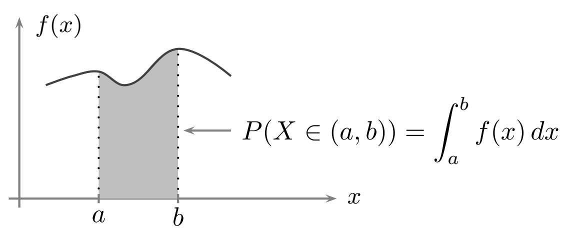


Figura 1.15: La probabilidad como un área.

Ejemplo. La función $f(x)$ dada por

$$f(x) = \begin{cases} 1/2 & \text{si } x \in (1,3), \\ 0 & \text{otro caso,} \end{cases}$$

es una función de densidad de una variable aleatoria continua que toma valores en el intervalo $(1,3)$, y cuya gráfica aparece en la Figura 1.16. Observe que se trata de una función no negativa y cuya integral vale uno.

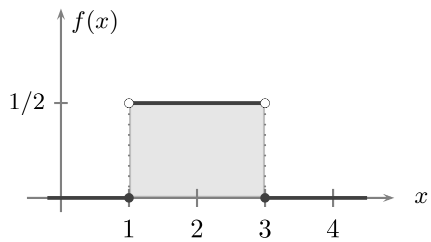


Figura 1.16:

▪

Ejemplo. Encontraremos el valor de la constante c que hace que la siguiente función sea de densidad.

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

Se trata de una variable aleatoria continua que toma valores en el intervalo $[-1,1]$. Como esta función debe integrar uno tenemos que

$$1 = \int_{-1}^{1} c\,|x|\,dx = 2\int_{0}^{1} c\,x\,dx = c.$$

Por lo tanto, cuando tomamos $c=1$ la función del inciso (b) resulta ser una función de densidad pues ahora cumple con ser no negativa e integrar uno. ▪

Ejercicio. Grafique y compruebe que las siguientes funciones son de densidad.

a) $f(x) = (x+1)/2$, para $x \in (-1,1)$.

b) $f(x) = e^{-x}$, para $x > 0$.

▪

Función de distribución. Sea X una variable aleatoria discreta o continua. La *función de distribución* de X, denotada por $F(x) : \mathbb{R} \to [0, 1]$, se define como $F(x) = P(X \leq x)$. Esto es, la función de distribución evaluada en un número x cualquiera es simplemente la probabilidad de que la variable aleatoria tome un valor menor o igual a x, o en otras palabras, que tome un valor en el intervalo $(-\infty, x]$. Siendo $F(x)$ una probabilidad, sus valores están siempre entre 0 y 1. Esta función resulta ser importante y se le conoce también, por razones evidentes, con el nombre de *función de acumulación de probabilidad*. Con un par de ejemplo mostraremos la forma de calcular esta función a partir de la función de probabilidad o de densidad.

Ejemplo. Considere la variable aleatoria discreta X de la Figura 1.14. Tenemos que la correspondiente función de distribución evaluada en x se calcula sumando las probabilidades $P(X = u)$ para valores de u menores o iguales a x, es decir,

$$F(x) = P(X \leq x) = \sum_{u \leq x} P(X = u) = \begin{cases} 0 & \text{si } x < 1, \\ 0.3 & \text{si } 1 \leq x < 2, \\ 0.8 & \text{si } 2 \leq x < 3, \\ 1 & \text{si } x \geq 3, \end{cases}$$

cuya gráfica aparece en la Figura 1.17 (a). Este es el comportamiento típico de una función de distribución discreta, es no decreciente, constante por pedazos, y si la función tiene una discontinuidad en x, entonces el tamaño de tal discontinuidad es exactamente la probabilidad de que la variable aleatoria tome ese valor. ▪

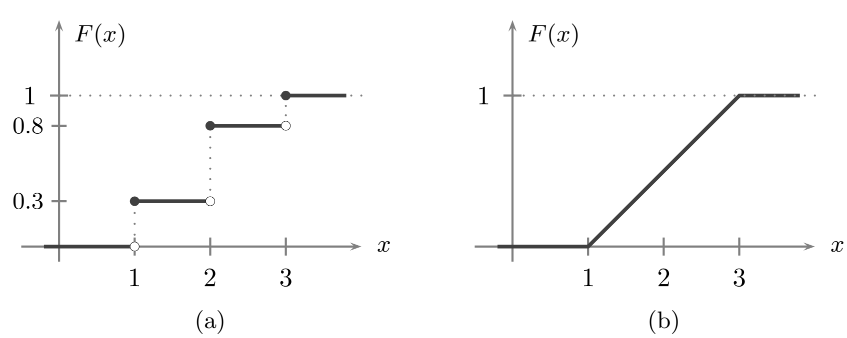


Figura 1.17:

Ejemplo. Considere ahora la variable aleatoria continua X de la Figura 1.17 (b). La correspondiente función de distribución se obtiene calculando la siguiente integral

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du = \begin{cases} 0 & \text{si } x \leq 1, \\ (x-1)/2 & \text{si } 1 < x < 3, \\ 1 & \text{si } x \geq 3, \end{cases}$$

cuya gráfica aparece en la Figura 1.17 (b). Observe que esta función es continua y no decreciente. ▪

En los dos ejemplos anteriores se ha mostrado la forma de obtener $F(x)$ a partir de $f(x)$. Ahora explicaremos el proceso contrario. En el caso continuo tenemos que para toda x en $\mathbb{R}$,

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

de modo que por el teorema fundamental del cálculo, y cuando $F(x)$ es diferenciable, $F'(x) = f(x)$. De este modo podemos encontrar $f(x)$ a partir de $F(x)$.

En el caso discreto, $f(x) = P(X = x) = F(x) - F(x-)$, en donde $F(x-)$ es el líimite por la izquierda de la función F en el punto x, en símbolos, $F(x-) = \lim_{h \to 0} F(x-h)$, con $h > 0$. Análogamente, la expresión $F(x+)$ significa el límite por la derecha de la función F en el punto x, es decir, $F(x+) = \lim_{h \to 0} F(x+h)$, con $h > 0$.

Proposición. Toda función de distribución $F(x)$ satisface las siguientes propiedades:

a) $0 \leq F(x) \leq 1$.

b) $\lim_{x \to \infty} F(x) = 1$.

c) $\lim_{x \to -\infty} F(x) = 0$.

d) Si $x_1 \leq x_2$, entonces $F(x_1) \leq F(x_2)$.

e) Si $x_1 \leq x_2$, entonces $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$.

f) $F(x) = F(x+)$.

Demostración.

a) Primeramente tenemos que $F(x)$ es una probabilidad pues, por definición, $F(x) = P(X \leq x)$. Por lo tanto se cumple la primera propiedad.

b) Cuando x tiende a infinito el conjunto $(X \leq x)$ se aproxima al conjunto $(X \leq \infty)$ que es idéntico a Ω, por lo tanto, cuando $x \to \infty$, $F(x) \to P(X \leq \infty) = P(\Omega) = 1$.

c) Análogamente el conjunto $(X \leq x)$ se aproxima al conjunto $(X \leq -\infty) = \emptyset$ cuando x tiende a menos infinito. Por lo tanto, cuando $x \to -\infty$, $F(x) \to P(X \leq -\infty) = P(\emptyset) = 0$.

d) Es suficiente observar que si $x_1 \leq x_2$, entonces $(X \leq x_1) \subseteq (X \leq x_2)$. Aplicando probabilidad obtenemos $P(X \leq x_1) \leq P(X \leq x_2)$.

e) Observe que el evento $(x_1 < X \leq x_2)$ puede descomponerse en la diferencia $(X \leq x_2) - (X \leq x_1)$, en donde $(X \leq x_1) \subseteq (X \leq x_2)$. Por lo tanto $P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$.

f) Para $h > 0$ tenemos que $F(x+h) = P(X \leq x+h) = P(X \leq x) + P(x < X \leq x+h)$, de modo que cuando h tiende a cero, el conjunto $(x < X \leq x+h)$ tiende al conjunto vacío. Concluimos entonces que, cuando $h \to 0$ con $h > 0$, $F(x+h) \to F(x) + P(\emptyset) = F(x)$.

□

La propiedad d) significa que $F(x)$ es una función monótona no decreciente. Mientras que la propiedad f) establece que $F(x)$ es una función continua por la derecha.

1.7. Esperanza, varianza, momentos

Todos los seres humanos tenemos características numéricas que nos identifican y nos distinguen de otras personas, por ejemplo, la edad, estatura, talla, peso, etc. Si pudiéramos considerar la totalidad de todos estos números para una persona en particular, la identificaríamos de manera única. Algo similar sucede con las variables aleatorias. En esta sección estudiaremos algunas características numéricas asociadas a las variables aleatorias.

Esperanza. La esperanza de una variable aleatoria X es un número denotado por $E(X)$ y que se calcula como sigue: Si X es discreta, entonces

$$\mathrm{E}(X) = \sum_x x P(X = x),$$

en donde la suma se efectúa sobre todos los posibles valores que pueda tomar la variable aleatoria, y se define cuando esta suma sea absolutamente convergente. El número de sumandos puede ser finito o infinito dependiendo del conjunto de valores de la variable aleatoria. Si X es continua con función de densidad $f(x)$, entonces la esperanza es

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

suponiendo que esta integral es absolutamente convergente. Si la suma o integral anteriores no cumplen esta condición de convergencia absoluta, entonces se dice que la esperanza no existe. La esperanza de una variable aleatoria es entonces un número que indica el promedio ponderado de los diferentes valores que puede tomar la variable. A la esperanza se le conoce también con los nombre de: *media*, *valor esperado* o *valor promedio*. En general se usa la letra griega μ (mu) para denotarla. La integral o suma arriba mencionados pueden no ser convergentes y en ese caso se dice que la variable aleatoria no tiene esperanza finita. La situación anterior se ilustra en los ejercicios 126 y 127. La esperanza es uno de los conceptos más importantes en probabilidad y tiene un amplio uso en las aplicaciones y otras ramas de la ciencia. Ilustraremos a continuación la forma de calcular la esperanza.

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

x	-1	0	1	2
$f(x)$	1/8	4/8	1/8	2/8

La esperanza de X es el número

$$\begin{aligned} E(X) &= \sum_x x\, f(x) \\ &= -1 \times 1/8 + 0 \times 4/8 + 1 \times 1/8 + 2 \times 2/8 \\ &= 1/2. \end{aligned}$$

Observe que la suma su efectúa para todos los valores de x indicados en la tabla, es decir: $-1, 0, 1$ y 2. También es instructivo observar que la esperanza no es necesariamente uno de los valores tomados por la variable aleatoria. En este ejemplo el valor 1/2 nunca es tomado por la variable aleatoria, pero es su valor esperado. ▪

Ejemplo. Considere la variable aleatoria continua X con función de densidad $f(x) = 2x$, para $x \in (0,1)$, siendo cero fuera de este intervalo. La esperanza de X es

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^1 x\, 2x\, dx = \frac{2}{3}x^3 \Big|_0^1 = \frac{2}{3}.$$

Observe que la integral sólo es relevante en el intervalo $(0,1)$, pues fuera de dicho intervalo la función de densidad se anula. ▪

Ejercicio. Considere una variable aleatoria continua X con función de densidad $f(x) = |x|$, para $-1 < x < 1$. Compruebe que la esperanza de esta variable aleatoria es cero. ▪

Ejercicio. Sea X una variable aleatoria discreta con posibles valores en el conjunto $\{1, 2, \ldots, n\}$, y tal que la probabilidad de que tome cualquiera de estos valores es $1/n$. Compruebe que la esperanza de X es $(n+1)/2$. ▪

Esperanza de una función de una variable aleatoria. En algunos casos es necesario saber calcular la esperanza de una función de una variable aleatoria. Por ejemplo, si X es una variable aleatoria, entonces es claro que $Y = X^2$ es una función de X y es también una variable aleatoria. Si quisiéramos calcular la esperanza de Y según la definición tendríamos que calcular $\int y\, f_Y(y)\, dy$, para lo cual se necesita encontrar primero la función de densidad de Y, y ello en general no es fácil. El siguiente resultado es muy útil y nos dice cómo calcular esta esperanza conociendo únicamente la función de densidad de X. A veces se le refiere como el *teorema del estadístico inconsciente*.

Proposición. Sea X una variable aleatoria continua y sea $g : \mathbb{R} \to \mathbb{R}$ una función tal que $g(X)$ es una variable con esperanza finita. Entonces

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx. \tag{1.3}$$

En general, la demostración de este resultado es complicada, asi es que la omitiremos y nos concentraremos en su uso y aplicación. El resultado esta enunciado en el caso continuo pero también es válido en el caso discreto, en lugar de la integral aparece una suma.

Ejemplo. Calcularemos $E(Y)$ en donde $Y = X^2$, y X es la variable aleatoria continua del ejemplo anterior, es decir, con función de densidad $f(x) = 2x$ para $x \in (0,1)$. Por la proposición anterior tenemos que

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

Ahora, como un ejercicio, encuentre la función de densidad de Y y calcule $E(Y)$ usando la definición de esperanza. El resultado debe ser nuevamente $1/2$. ▪

Ejercicio. Sea X una variable aleatoria con función de probabilidad dada por la tabla que aparece abajo. Encuentre la función de probabilidad de X^2 y mediante la definición calcule $E(X^2)$. Ahora calcule la misma esperanza pero usando (1.3). Ambos resultados deben coincidir.

x	-2	-1	0	1	2
$f(x)$	2/8	1/8	2/8	1/8	2/8

▪

He aqui algunas propiedades generales de la esperanza.

Proposición. Sean X y Y con esperanza finita y sea c una constante. Entonces

a) $E(c) = c$.

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

c) Si $X \geq 0$, entonces $E(X) \geq 0$.

d) $E(X+Y) = E(X) + E(Y)$.

La primera propiedad es evidente de demostrar pues si X es la variable aleatoria constante c, entonces por definición, $E(X) = c\,P(X = c) = c\,1 = c$. El segundo inciso se sigue directamente de la definición de esperanza pues tanto en el caso de la suma como en el caso de la integral, la constante c puede siempre colocarse fuera. El tercer inciso también es evidente pues cuando se cumple la hipótesis, en la integral o suma correspondiente solo aparecerán términos que son no negativos. La última propiedad, en cambio, no es sencilla de demostrar y aún en el caso discreto requiere de detalles técnicos que preferimos omitir. Observe que la segunda y cuarta propiedad establecen que la esperanza es *lineal*, es decir, separa sumas y también separa multiplicaciones por constantes.

Varianza. Vamos ahora a definir otra característica numérica asociada a las variables aleatorias llamada *varianza*. Se denota por $\text{Var}(X)$ y se define como sigue.

$$\text{Var}(X) = \begin{cases} \displaystyle\sum_x (x - E(X))^2 f(x) & \text{si } X \text{ es discreta.} \\ \displaystyle\int_{-\infty}^{\infty} (x - E(X))^2\, f(x)\, dx & \text{si } X \text{ es continua.} \end{cases}$$

Observe que en una sola expresión la varianza se puede escribir como sigue: $\text{Var}(X) = E\left[(X - E(X))^2\right]$. La varianza es una medida del grado de dispersión de los diferentes valores tomados por la variable. Se le denota regularmente por la letra σ^2 (sigma cuadrada). A la raíz cuadrada positiva de la varianza, esto es σ, se le llama *desviación estándar*. Nuevamente la anterior suma o integral puede no existir y en ese caso decimos que la variable aleatoria no tiene varianza finita. Observemos que para calcular $\text{Var}(X)$ necesitamos conocer primero $E(X)$. Veamos algunos ejemplos sencillos.

Ejemplo. Calcularemos la varianza de la variable aleatoria discreta X con función de densidad dada por la siguiente tabla.

x	-1	0	1	2
$f(x)$	1/8	4/8	1/8	2/8

Recordemos primeramente que por cálculos previos, $E(X) = 1/2$. Aplicando la definición de varianza tenemos que

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_x (x - E(X))^2 f(x) \\
&= (-1 - 1/2)^2 \times 1/8 + (0 - 1/2)^2 \times 4/8 \\
&\quad + (1 - 1/2)^2 \times 1/8 + (2 - 1/2)^2 \times 2/8 \\
&= 1.
\end{aligned}$$

▪

Ejemplo. Calcularemos la varianza de la variable aleatoria continua X con función de densidad $f(x) = 2x$ para $x \in (0,1)$. En un cálculo previo habíamos encontrado que $E(X) = 2/3$. Por lo tanto,

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\, dx \\
&= \int_0^1 (x - 2/3)^2\, 2x\, dx \\
&= \int_0^1 (2x^3 - 8x^2/3 + 8x/9)\, dx \\
&= (x^4/2 - 8x^3/9 + 4x^2/9)\Big|_0^1 \\
&= 1/18.
\end{aligned}$$

▪

Ahora enunciamos algunas propiedades de la varianza.

Proposición. Sean X y Y dos variables aleatorias, y sea c una constante. Entonces

a) $\mathrm{Var}(X) \geq 0$.

b) $\mathrm{Var}(c) = 0$.

c) $\mathrm{Var}(c\,X) = c^2\,\mathrm{Var}(X)$.

d) $\mathrm{Var}(X + c) = \mathrm{Var}(X)$.

e) $\mathrm{Var}(X) = E(X^2) - E^2(X)$.

f) En general, $\mathrm{Var}(X + Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Demostración. El inciso (a) es evidente a partir de la definición de varianza pues en ella aparece una suma o integral de términos no negativos. Para el inciso (b) la constante c es una v.a. con un único valor, de modo que $E(c) = c$ y entonces $\mathrm{Var}(X) = E(c - c)^2 = 0$. Para el inciso ($c$) tenemos que

$$\begin{aligned}\mathrm{Var}(cX) &= E[cX - E(cX)]^2\\ &= E[cX - cE(X)]^2\\ &= c^2\,E[X - E(X)]^2\\ &= c^2\,\mathrm{Var}(X).\end{aligned}$$

El inciso (d) se sigue del siguiente análisis:

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

Para demostrar la propiedad (e) se desarrolla el cuadrado en la definición de varianza, y se usa la propiedad de linealidad de la esperanza:

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

Finalmente para demostrar la propiedad (f) es suficiente dar un ejemplo. Puede tomarse el caso $Y = X$, en general y por lo demostrado antes, no se cumple que $\mathrm{Var}(2X) = 2\,\mathrm{Var}(X)$. $\square$

Momentos. Finalmente definimos el *n-ésimo momento* de una variable aleatoria X, cuando existe, como el número $E(X^n)$, para cualquier valor natural de n. El *n-ésimo momento central* de X, cuando existe, es el número $E[(X-\mu)^n]$, en donde $\mu = E(X)$. Observe que el primer momento de X es simplemente la media, y el segundo momento central es la varianza. Tenemos entonces que si X es una variable aleatoria con función de densidad o de probabilidad $f(x)$ entonces el n-ésimo momento de X, si existe, se calcula como sigue:

$$E(X^n) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x^n\, f(x)\, dx, \\ \displaystyle\sum_x x^n\, f(x)\,. \end{cases}$$

El n-ésimo momento central de X se calcula, para variables aleatorias continuas y discretas respectivamente, como indican las siguientes fórmulas:

$$E[(X-\mu)^n] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} (x-\mu)^n\, f(x)\, dx, \\ \displaystyle\sum_x (x-\mu)^n\, f(x)\,. \end{cases}$$

1.8. Distribuciones de probabilidad

Estudiaremos a continuación algunas distribuciones de probabilidad de variables aleatorias importantes. Estas distribuciones son modelos particulares para asignar probabilidades a subconjuntos de números reales. Empezaremos con las distribuciones de tipo discreto y continuaremos después con las de tipo continuo. Es importante señalar que ésta es sólamente una lista parcial de algunas distribuciones de probabilidad de mayor uso.

Distribución uniforme discreta. Decimos que una variable aleatoria X tiene una distribución uniforme discreta sobre el conjunto finito de números $\{x_1, \ldots, x_n\}$

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 ocurrir. 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

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

Por ejemplo, la gráfica de la función de probabilidad de la distribución $\text{unif}\{1, 2, 3, 4, 5\}$ aparece en la Figura 1.18, junto con la correspondiente función de distribución. Cada salto de la función de distribución es de tamaño $1/5$. Es fácil ver que $E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$, y $\text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))^2$.

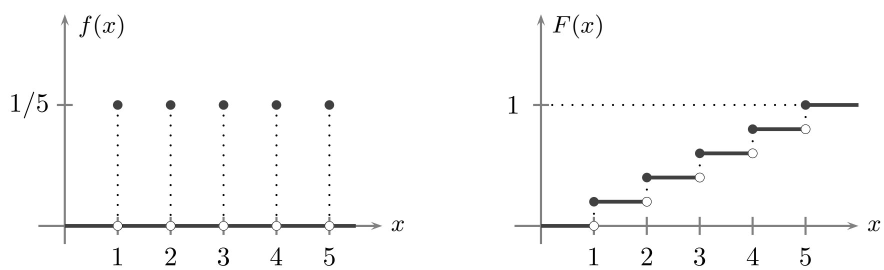


Figura 1.18:

Ejemplo. Al generar un número aleatorio en una computadora dentro del intervalo unitario $[0, 1]$, y debido a que la precisión de la computadora es necesariamente finita, se obtienen siempre valores dentro de un conjunto finito de elementos. Por ejemplo, si la precisión de la computadora fuera de dos decimales, entonces sólo se pueden generar los 101 números : 0.00, 0.01, 0.02,. . ., 0.99, 1.00. La precisión de una computadora es naturalmente mayor pero siempre es finita, y la misma situación se presenta. ▪

Distribución Bernoulli. Un *ensayo Bernoulli* se define como aquel experimento aleatorio con únicamente dos posibles resultados, llamados genéricamente éxito y fracaso, con probabilidades respectivas p y $1 - p$. Si se define la variable aleatoria X como aquella función que lleva el resultado éxito al número 1 y el resultado fracaso al número 0, entonces decimos que X tiene una distribución Bernoulli con parámetro $p \in (0, 1)$, y escribimos $X \sim \text{Ber}(p)$. La función de probabilidad es entonces

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

La gráfica de la función de densidad de esta distribución para p =0.7 aparece en la Figura 1.19, junto con la correspondiente función de distribución. En este caso es muy sencillo verificar que $E(X) = p$ y $\text{Var}(X) = p(1-p)$. En la realización de todo experimento aleatorio siempre es posible preguntarnos por la ocurrencia o no ocurrencia de un evento cualquiera. Por ejemplo, ganar o no ganar en un juego de lotería, que llueva o no llueva hoy por la tarde, etc. Este es el esquema general donde surge esta distribución, que aunque sencilla, es de amplia aplicación.

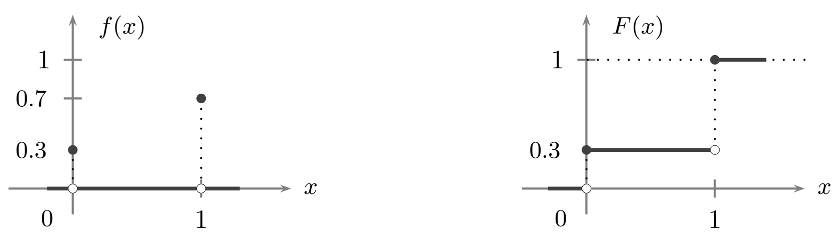


Figura 1.19:

Ejemplo. Considere el experimento aleatorio de lanzar una moneda al aire. Suponga que ω_1 y ω_2 son los dos resultados posibles, con probabilidades p y $1-p$, respectivamente. Sea X la variable aleatoria dada por $X(\omega_1) = 1$, y $X(\omega_2) = 0$. Entonces X tiene distribución Ber(p). ¿Cuál es la distribución de $1 - X$? ▪

Ejemplo. Sea Ω el espacio muestral de un experimento aleatorio, y sea $A \subseteq \Omega$ un evento tal que $P(A) = p$. Sea X la variable aleatoria dada por $X(\omega) = 1_A(\omega)$, es decir, $X(\omega) = 1$ si $\omega \in A$, y $X(\omega) = 0$ si $\omega \in A^c$. Esta es la *función indicadora* del conjunto A. Entonces X tiene distribución Ber(p).

▪

Distribución binomial. Supongamos ahora que tenemos una serie de n ensayos independientes Bernoulli en donde la probabilidad de éxito en cualesquiera de estos ensayos es p. Si denotamos por E el resultado éxito y por F el resultado fracaso, entonces el espacio muestral consiste de todas las posibles sucesiones de longitud n de caracteres E y F. Usando el principio multiplicativo, es fácil ver que 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, $X(EE\cdots EE) = n$, $X(FE\cdots EE) = n-1$, ..., $X(FF\cdots FF) = 0$, entonces tenemos que X puede tomar los valores $0, 1, 2, \ldots, n$ con las probabilidades que aparecen abajo. Decimos entonces que X tiene una distribución binomial con parámetros n y p, y escribimos $X \sim \text{bin}(n, p)$.

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

Por ejemplo, para $n = 10$ ensayos, con probabilidad p =0.3, se puede calcular $P(X = 2) = \binom{10}{2}\,(0.3)^2\,(0.7)^{10-2} = 0.2334$. La gráfica de esta función de probabilidad con estos parámetros aparece en la Figura 1.20. 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

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

pero hemos colocado los x éxitos en los primeros x ensayos, tenemos entonces que multiplicar por las diferentes formas en que estos x éxitos pueden distribuirse en los n ensayos, este factor es el coeficiente binomial $\binom{n}{x}$. Para esta distribución puede demostrarse que $E(X) = np$, y $\text{Var}(X) = np(1-p)$.

Distribución geométrica. Supongamos que tenemos ahora una sucesión infinita de ensayos independientes Bernoulli, en cada uno de los cuales la probabilidad de éxito es p. Para cada una de estas sucesiones definimos la variable aleatoria X como el número de fracasos antes de obtener el primer éxito. Por ejemplo, $X(FEFEFF\cdots) = 1$, $X(EFFEEE\cdots) = 0$, $X(FFFEFE\cdots) = 3$. Observamos que X puede tomar los valores $0, 1, 2, \ldots$ La probabilidad de que X tome el valor entero $x \geq 0$ es $p(1-p)^x$. Decimos entonces que X tiene una distribución geométrica con parámetro p, y escribimos $X \sim \text{geo}(p)$ cuando

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

Por ejemplo, para p =0.4, la gráfica de esta función se muestra en la Figura 1.21. El nombre de esta distribución proviene del hecho de que cuando escribimos la

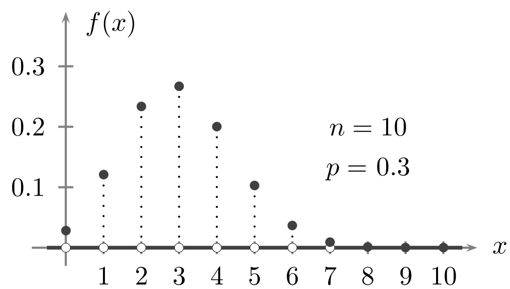


Figura 1.20:

suma de todas las probabilidades, obtenemos una suma geométrica. La inspección sucesiva de artículos hasta encontrar una defectuoso, posiblemente en un proceso de control de calidad, puede modelarse usando una distribución geométrica. Para esta distribución tenemos que $E(X) = (1-p)/p$, y $\text{Var}(X) = (1-p)/p^2$.

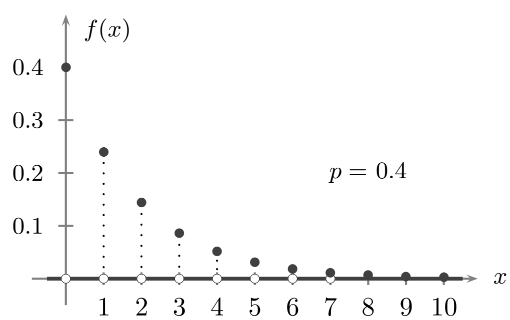


Figura 1.21:

En algunos textos se define la distribución geométrica contando el número de ensayos (no el de fracasos), antes del primer éxito. En este caso la variable es $X + 1$, es decir, la distribución se desplaza hacia la derecha una unidad. la esperanza es ahora $1/p$ y la varianza permanece constante $(1-p)/p^2$.

Ejercicio. Una persona participa cada semana con un boleto en un juego de lotería en donde la probabilidad de ganar el primer premio es $p = 10^{-6} = 1/1,000,000$.

¿Cuántos años en promedio debe esta persona participar en el juego antes de obtener el primer premio? ▪

Ejercicio. Sea X una variable aleatoria con distribución geo(p). Demuestre que la función de distribución de X es

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1-(1-p)^1 & \text{si } 0 \leq x < 1, \\ 1-(1-p)^2 & \text{si } 1 \leq x < 2, \\ \cdots & \cdots \\ 1-(1-p)^{k+1} & \text{si } k \leq x < k+1, \\ \cdots & \cdots \end{cases}$$

▪

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 día. Para modelar este tipo de situaciones podemos definir la variable aleatoria X como el número de ocurrencia de este evento 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 (lambda). 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 entero $x \geq 0$ se definirá a continuación. Decimos que X tiene una distribución Poisson con parámetro $\lambda > 0$, y escribimos $X \sim \text{Poisson}(\lambda)$ cuando

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

Puede demostrarse que la función $f(x)$ arriba definida es efectivamente una función de probabilidad para cada valor de $\lambda > 0$. La forma de esta función se muestra en la Figura 1.22 cuando $\lambda = 2$. Después de algunos cálculos sencillo puede también comprobarse que $E(X) = \lambda$, y $\text{Var}(X) = \lambda$.

Ejemplo. En promedio se reciben 2 peticiones de acceso a una página web durante

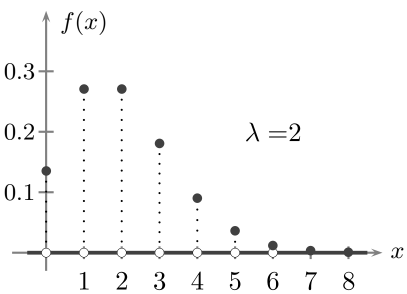


Figura 1.22:

un minuto cualquiera. Utilice el modelo Poisson para calcular la probabilidad de que en un minuto dado a) nadie solicite acceso a la página. b) se reciban mas de dos peticiones. Solución: Sea X el número de peticiones por minuto, y supongamos que X tiene distribución Poisson(λ), con $\lambda = 2$. Para el primer inciso se tiene que $P(X = 0) = e^{-2}\, 2^0/0! =$0.135. Para el segundo inciso, $P(X > 2) = 1 - P(X \leq 2)$ $= 1-(P(X = 0)+P(X = 1)+P(X = 2)) = 1-e^{-2}\,(2^0/0!+2^1/1!+2^2/2!) =$0.323 ■

Puede además 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.

Este resultado sugiere que cuando n es grande, la distribución binomial puede ser aproximada mediante la distribución Poisson de parámetro $\lambda = np$. Esto es particularmente útil pues el cálculo de probabilidades de la distribución binomial involucra el cálculo de factoriales y ello puede ser computacionalmente difícil. Un ejemplo ilustrará esta situación.

Ejemplo. En promedio uno de cada 100 focos producido por una máquina es defectuoso. Calcule la probabilidad de encontrar 5 focos defectuosos en un lote de 1000 focos. Sugerencia: Use la distribución Poisson como aproximación de la distribución binomial. ■

Hemos definido a la variable aleatoria Poisson como el número de ocurrencias de un cierto evento dentro de un intervalo de tiempo dado, supongamos de longitud unitaria, $[0,1]$. Suponga ahora que nos interesa observar las ocurrencias del evento en un intervalo de longitud diferente, por ejemplo $[0,t]$, con $t>0$. Tal conteo de ocurrencias también sigue una distribución Poisson pero esta vez de parámetro λt. Por ejemplo, si $t=2$, entonces el número de ocurrencias del evento en el intervalo $[0,2]$ tiene distribución Poisson(λt). Veamos un ejemplo.

Ejemplo. El número de aviones que llegan a un aeropuerto internacional tiene una distribución Poisson con una frecuencia de 3 aviones cada 10 minutos. Entonces a) La probabilidad de que no llegue ningún avión en un periodo de 20 minutos es $P(X=0)$, con $\lambda=6$. b) La probabilidad de que llegue sólo un avión en el minuto siguiente es $P(X=1)$, con $\lambda=3/10$. c) La probabilidad de que lleguen dos o mas aviones en un periodo de 15 minutos es $P(X\geq 2)$, con λ =4.5. ▪

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 r y p, y escribimos $X\sim \text{bin neg}(r,p)$. En este caso tenemos que X puede tomar los valores $0,1,2,\ldots$ con probabilidades como se indica a continuación.

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

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 ahí el término $(1-p)^x$, y finalmente el factor $\binom{r+x-1}{x}$ 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. La gráfica de esta función aparece en la Figura 1.23 cuando $r=3$ y p =0.2.

Es claro que esta distribución es una generalización de la distribución geométrica, la cual se obtiene tomando $r=1$. Se puede además demostrar que $E(X)=r(1-p)/p$ y $\text{Var}(X)=r(1-p)/p^2$.

La definición de coeficiente binomial puede extenderse del siguiente modo: Para cualquier número real a, y cualquier entero natural x se define $\binom{a}{x}=a(a-1)\cdots(a-x+1)/x!$ Puede demostrarse la identidad $\binom{n+x-1}{x}=(-1)^x\binom{-n}{x}$. De este hecho adquiere su nombre esta distribución.

Ejemplo. Se lanza repetidas veces una moneda honesta, cuyos dos resultados son

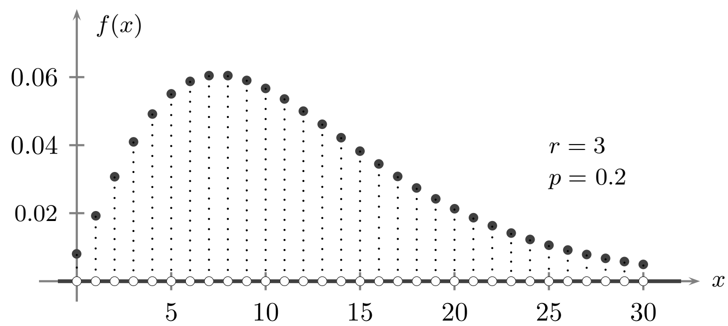


Figura 1.23:

cara y cruz. ¿Cuál es la probabilidad de obtener la tercera cruz en el quinto lanzamiento? Solución: Sea X el número de caras (fracasos) antes de obtener la tercera cruz. Entonces $X \sim \text{bin neg}(r, p)$ con $r = 3$, $p = 1/2$, y nos preguntan $P(X = 2)$. Por lo tanto, $P(X = 2)= \binom{6}{2}(1/2)^5= 15/32= 0.46875$.

Distribución hipergeométrica. Supongamos que tenemos un conjunto de N objetos de los cuales K son de una primera clase y $N - K$ son de una segunda clase. Supongamos que de este conjunto tomamos una muestra aleatoria de tamaño n, la muestra es sin reemplazo y el orden de los objetos seleccionados no importa. El espacio muestral de este experimento consiste de todas las posibles muestras de tamaño n que se pueden obtener del conjunto mayor de tamaño N. La cardinalidad del espacio muestral es $\binom{N}{n}$. 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 con parámetros N, K y n, y escribimos $X \sim \text{hipergeo}(N, K, n)$ si

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

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 $\binom{N-K}{n-x}$ es nuevamente las diferentes formas de escoger $n-x$ objetos de la totalidad de $N-K$ 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. La gráfica de esta función de densidad para ciertos valores de los parámetros aparece en la Figura 1.24. En este caso es posible comprobar que $E(X) = nK/N$, y $\text{Var}(X) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$.

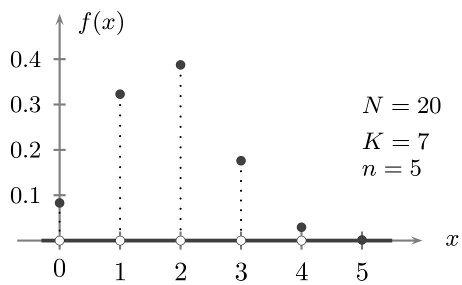


Figura 1.24:

Con esto concluimos la revisión de algunas distribuciones de probabilidad de tipo discreto. Presentamos en la siguiente tabla un resumen de las distribuciones discretas mencionadas en esta sección.

Distribución uniforme

$f(x) = 1/n$
para $x = x_1, \ldots, x_n$.
Parámetros: $x_1, \ldots, x_n; n$.
Media: $\sum_{i=1}^n x_i/n$.
Varianza: $\sum_{i=1}^n (x_i - \mu)^2/n$.

Distribución Bernoulli

$f(x) = p^x\,(1-p)^{1-x}$
para $x = 0, 1$.
Parámetro: $p \in (0,1)$.
Media: p.
Varianza: $p(1-p)$.

Distribución binomial

$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
para $x = 0, 1, \ldots, n$.
Parámetros: $n \in \mathbb{N}$, $p \in (0,1)$.
Media: np.
Varianza: $np(1-p)$.

Distribución geométrica

$f(x) = p\,(1-p)^x$
para $x = 0, 1, 2, \ldots$
Parámetro: $p \in (0,1)$.
Media: $(1-p)/p$.
Varianza: $(1-p)/p^2$.

Distribución Poisson

$f(x) = \frac{\lambda^x}{x!}\, e^{-\lambda}$
para $x = 0, 1, 2, \ldots$
Parámetro: $\lambda > 0$.
Media: λ.
Varianza: λ.

Distribución binomial negativa

$f(x) = \binom{r+x-1}{x} p^r (1-p)^x$
para $x = 0, 1, 2, \ldots$
Parámetros: $r \in \mathbb{N}$, $p \in (0,1)$.
Media: $r(1-p)/p$.
Varianza: $r(1-p)/p^2$.

Distribución hipergeométrica

$f(x) = \binom{K}{x}\binom{N-K}{n-x}/\binom{N}{n}$
para $x = 0, 1, \ldots, n$.
Parámetros: N, K, n.
Media: nK/N.
Varianza: $nK(N-K)(N-n)/(N^2(N-1))$.

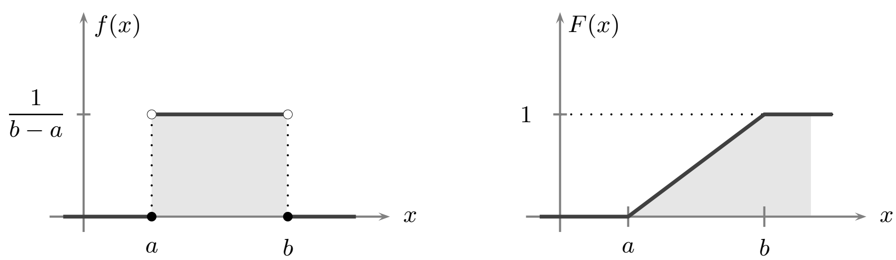
Ahora estudiaremos algunas distribuciones de probabilidad de variables aleatorias continuas. No construiremos estas distribuciones a partir de experimentos aleatorios particulares como en el caso de las distribuciones discretas, mas bien, las definiremos sin mayor justificación, en algunos casos mostraremos la forma de obtener estas distribuciones a partir de considerar ciertas funciones de variables aleatorias conocidas.

Distribución uniforme continua. Decimos que una variable aleatoria X tiene una distribución uniforme continua en el intervalo (a,b), y escribimos $X \sim$

unif(a, b), cuando su función de densidad es

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

La gráfica general de esta función se muestra en la Figura 1.25(a), y es evidente que se trata de una función de densidad pues es no negativa e integra uno. En este caso es muy fácil encontrar la correspondiente función de distribución, y ésta se muestra en la Figura 1.25(b). Los parámetros de esta distribución son los números $a < b$. Es fácil verificar que $E(X) = (a+b)/2$, que corresponde al punto medio del intervalo (a, b). Además $\text{Var}(X) = (b-a)^2/12$, de modo que la varianza o dispersión crece cuando a y b se alejan uno del otro, y por el contrario, cuando los parámetros estan muy cercanos, la varianza es pequeña. Esta distribución es una de las más sencillas, y sea usa naturalmente para cuando no se conoce mayor información de la variable aleatoria de interés, excepto que toma valores continuos dentro de algún intervalo.



(a) (b)

Figura 1.25:

Ejemplo. En el experimento aleatorio teórico de generar un número al azar X en el intervalo unitario $(0, 1)$ se considera regularmente que X tiene distribución uniforme en dicho intervalo. ▪

Ejercicio. Compruebe que la función de distribución de una variable aleatoria con distribución unif(a, b) es la que aparece en la Figura 1.25(b), y está dada por

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

▪

Distribución exponencial. Decimos que una variable aleatoria continua X tiene una distribución exponencial con parámetro $\lambda > 0$, y escribimos $X \sim \exp(\lambda)$, cuando su función de densidad es

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

La gráfica de esta función cuando el parámetro λ toma el valor particular 3, es la que se muestra en la Figura 1.26(a). La correspondiente función de distribución aparece a su derecha. Es muy sencillo verificar que la función $f(x)$ arriba definida es efectivamente una función de densidad para cualquier valor del parámetro $\lambda > 0$. Aplicando el método de integración por partes puede también comprobarse que $E(X) = 1/\lambda$, y $\text{Var}(X) = 1/\lambda^2$. Esta distribución se ha usado para modelar tiempos de espera para la ocurrencia de un cierto evento.

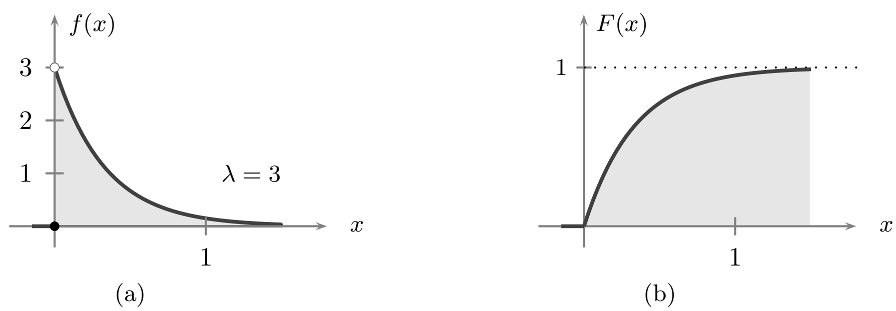


Figura 1.26:

Ejemplo. Suponga que el tiempo en minutos que un usuario cualquiera permanece revisando su correo electrónico sigue una distribución exponencial de parámetro $\lambda = 1/5$. Calcule la probabilidad de que un usuario cualquiera permanezca conectado al servidor de correo a) menos de un minuto. b) mas de un ahora.
Solución: Para el primer inciso tenemos que $P(X < 1) = \int_0^1 1/5\, e^{-x/5}\, dx$=0.181.
Para el segundo inciso, $P(X > 60)$= $\int_{60}^\infty 1/5\, e^{-x/5}\, dx$=0.0000061. ▪

Ejercicio. Sea X una variable aleatoria con distribución $\exp(\lambda)$. Compruebe que la función de distribución de X es la siguiente. Véase la Figura 1.26(b).

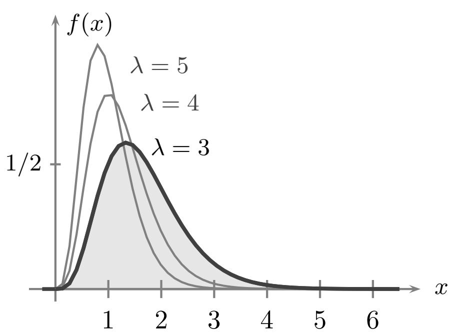
$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

▪

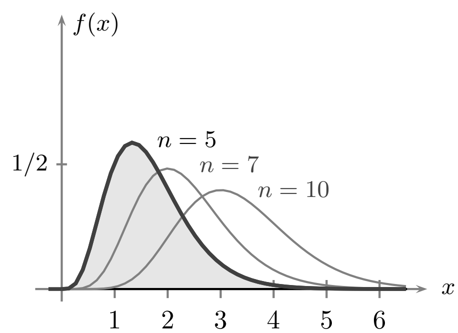
DISTRIBUCIÓN GAMA. La variable aleatoria continua X tiene una distribución gama con parámetros $n > 0$ y $\lambda > 0$, y escribimos $X \sim \text{gama}(n, \lambda)$, si su función de densidad es

$$f(x) = \begin{cases} \dfrac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

La gráfica de esta función de densidad para varios valores de los parámetros se muestra en la Figura 1.27.



(a) $n = 5$



(b) $\lambda = 3$

Figura 1.27:

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

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

para cualquier número real n tal que esta integral sea convergente. Esta función no es el tipo de funciones a las que estamos acostumbrados en los cursos de matemáticas en donde regularmente se conoce la expresión exacta de una cierta función y se utiliza esta expresión para evaluarla. En este caso, para evaluar la función gama es necesario substituir el valor de n en el integrando y efectuar la integral infinita. Afortunadamente no evaluaremos esta integral para cualquier valor de n, sólo para algunos pocos valores, principalmente enteros, y nos ayudaremos de las siguientes propiedades:

a) $\Gamma(n+1) = n\,\Gamma(n)$.

b) $\Gamma(n+1) = n!$ si n es entero.

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

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

El nombre de esta distribución se deriva del hecho de que en su definición aparece la función gama. Observemos además que la distribución exponencial es un caso particular de la distribución gama. En efecto, si en la distribución gama tomamos el parámetro n igual a 1, obtenemos la distribución exponencial con parámetro λ. Resolviendo un par de integrales se puede demostrar que $E(X) = n/\lambda$, y $\mathrm{Var}(X) = n/\lambda^2$.

Distribución beta. Decimos que la variable aleatoria continua X tiene una distribución beta con parámetros $a > 0$ y $b > 0$, y escribimos $X \sim \mathrm{beta}(a,b)$, cuando su función de densidad es

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

El término $B(a,b)$ se conoce como la *función beta*, y de allí adquiere su nombre esta distribución. La función beta se define como sigue

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

para números reales $a > 0$ y $b > 0$. Esta función está relacionada con la función gama a través de la identidad

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Véase la sección de ejercicios para una lista de propiedades de esta función. Para la distribución beta(a,b) se tiene que $E(X) = a/(a+b)$, y $\text{Var}(X) = ab/((a+b+1)(a+b)^2)$.

Distribución normal. Esta es posiblemente la distribución de probabilidad de mayor importancia. Decimos que la variable aleatoria continua X tiene una distribución normal si su función de densidad está dada por la siguiente expresión

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2},$$

en donde $\mu \in \mathbb{R}$ y $\sigma > 0$ son dos parámetros. Escribimos entonces $X \sim \text{N}(\mu, \sigma^2)$. La gráfica de esta función de densidad tiene forma de campana como se puede apreciar en la Figura 1.28, en donde se muestra además el significado geométrico de los dos parḿetros.

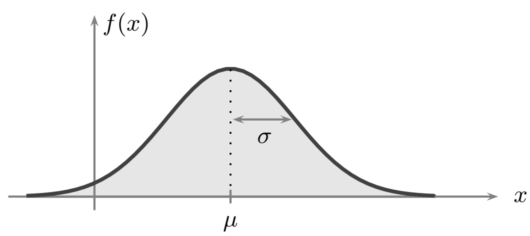


Figura 1.28:

No es inmediato pero es posible demostrar que $E(X) = \mu$, y ello significa que la campana esta centrada en este valor, el cual puede ser negativo, positivo o cero. También puede demostrarse que $\text{Var}(X) = \sigma^2$, y que la distancia del punto μ a cualquiera de los dos puntos en donde la función tiene puntos de inflexión es σ, por lo tanto la campana se abre o se cierra de acuerdo a la magnitud de este parámetro. En particular, decimos que la variable aleatoria 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

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}.$$

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

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

$$Z = \frac{X - \mu}{\sigma}. \tag{1.4}$$

A la operación anterior se le conoce con el nombre de *estandarización*, y bajo tal transformación se dice que la variable X ha sido *estandarizada*. Es común usar la letra Z para denotar una variable aleatoria con distribución normal estándar, y seguiremos nosotros también esa costumbre. Se deja como ejercicio al lector demostrar la proposición anterior, y si ello no es posible por ahora, por lo menos compruebe mentalmente las siguientes dos propiedades.

Ejercicio. A partir de (1.4) compruebe que $E(Z) = 0$ y $\text{Var}(Z) = 1$. ▪

La proposición anterior parece muy modesta pero tiene una gran importancia operacional pues establece que el cálculo de las probabilidades de una variable aleatoria normal cualquiera se reduce al cálculo de las probabilidades para la normal estándar. Explicaremos esta situación con más detalle. Suponga que X es una variable aleatoria con distribución $\text{N}(\mu, \sigma^2)$, y que deseamos calcular, por ejemplo, $P(a < X < b)$, para $a < b$ números dados. Tenemos entonces que

$$\begin{aligned} P(\,a < X < b\,) &= P(\,a - \mu < X - \mu < b - \mu\,) \\ &= P(\,\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\,) \\ &= P(\,\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\,). \end{aligned}$$

La igualdad de estas probabilidades es consecuencia de la igualdad de los eventos. De esta forma una probabilidad que involucra a la variable X se ha reducido a una probabilidad que involucra a Z. Es también común denotar la función de distribución de Z como $\Phi(x)$, es decir,

$$\Phi(x) = P(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-u^2/2}\, du,$$

cuyo significado geométrico se muestra en la Figura 1.29(a). Es necesario decir que, sin importar el método de integración que se utilice, no es posible encontrar una expresión exacta para esta integral, pero usando métodos numéricos pueden encontrarse aproximaciones de ella para distintos valores de x. En una tabla al final del texto aparece una tabla con estos valores aproximados. Cada renglón de esta tabla corresponde a un valor de x hasta el primer dígito decimal, las distintas columnas corresponden al segundo dígito decimal. El valor que aparece en la tabla es $\Phi(x)$. Por ejemplo, el renglón marcado con 1.4 y la columna marcada con 0.05 corresponden al valor 1.45, tenemos entonces que $\Phi(1.45)$ =0.9265. Aqui estan algunas preguntas para el lector: ¿Por qué en la tabla aparecen solo valores de x hasta 3.49?, es decir, ¿Cuánto vale $\Phi(x)$ para $x > 3.49$? ¿Cómo se podrían calcular los valores de $\Phi(x)$ para x negativos?

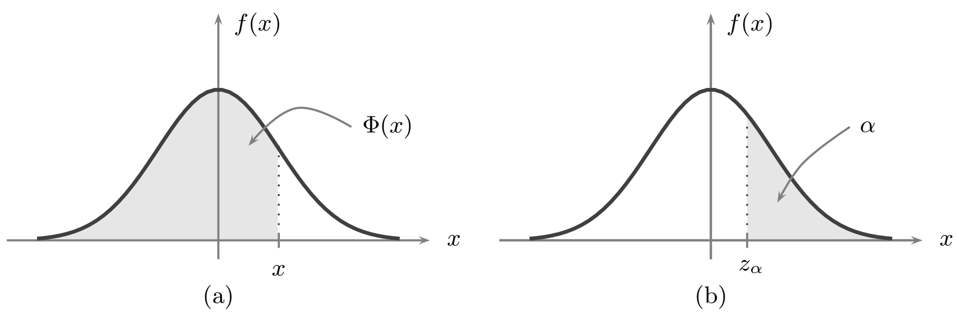


Figura 1.29:

Ejercicio. Demuestre que $\Phi(-x) = 1 - \Phi(x)$. ∎

Ejercicio. Encuentre x tal que a) $\Phi(x) = 0.3$. b) $\Phi(x) = 0.75$. ∎

Ejercicio. Use la tabla de la distribución normal estándar para comprobar que
a) $\Phi(1.65) = 0.9505$ b) $\Phi(-1.65) = 0.0495$ c) $\Phi(-1) = 0.1587$ ∎

Ejercicio. Sea X con distribución $\mathrm{N}(5, 10)$. Calcule
a) $P(X \leq 7)$. b) $P(0 < X < 5)$. c) $P(X > 10)$. ∎

NOTACIÓN z_α. Para cada valor α en el intervalo $(0, 1)$, el número z_α denotará el punto en el eje real para el cual el área bajo la curva a la derecha de z_α es α.

Esto es, $\Phi(z_\alpha) = 1 - \alpha$. Se muestra en la siguiente Figura 1.29(b) el significado geométrico del número z_α. Usaremos la notación z_α en la segunda parte de nuestro curso.

Acerca de la distribución normal tenemos el siguiente resultado de suma importancia.

Teorema central del límite. Sea $X_1, X_2, \ldots$ una sucesión infinita de variables aleatorias independientes e idénticamente distribuidas con media μ y varianza finita σ^2. Entonces la función de distribución de la variable

$$Z_n = \frac{(X_1 + \cdots + X_n) - n\mu}{\sqrt{n\sigma^2}}$$

tiende a la distribución normal estándar cuando n tiende a infinito, sin importar la distribución de cada variable de la sucesión, es decir,

$$\lim_{n\to\infty} F_{Z_n}(x) = F_Z(x).$$

Distribución ji-cuadrada. Decimos que la variable aleatoria continua X tiene una distribución ji-cuadrada con n grados de libertad (n entero positivo), si su función de densidad está dada por la siguiente expresión

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(n/2)} \left(\dfrac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Se trata de una variable aleatoria continua con posibles valores en el intervalo $(0, \infty)$. Esta distribución tiene un solo parámetro denotado aqui por la letra n, y al cual se le llama *grados de libertad*. A pesar de su aparente expresión complicada, no es difícil comprobar que $f(x)$ es efectivamente una función de densidad. La gráfica de esta función para varios valores del parámetro n aparece en la Figura 1.30. Escribiremos simplemente $X \sim \chi^2(n)$, en donde la letra griega χ se pronuncia "ji" o también "chi". Puede demostrarse que $E(X) = n$ y $\mathrm{Var}(X) = 2n$. La distribución ji-cuadrada puede obtenerse como indican los siguientes resultados que dejaremos sin demostrar.

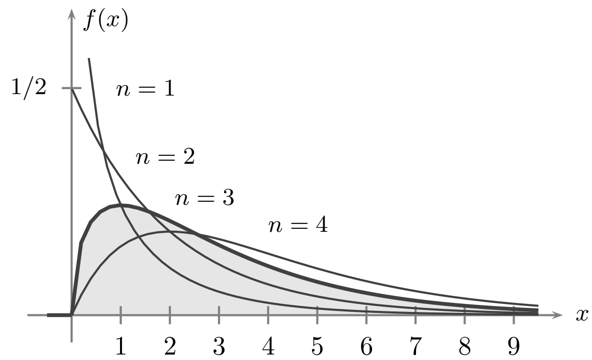


Figura 1.30:

Proposición. Si $X \sim \mathrm{N}(0,1)$, entonces $X^2 \sim \chi^2(1)$.

Es decir, el cuadrado de una variable aleatoria con distribución normal estándar tiene distribución ji-cuadrada con un grado de libertad. Por otro lado, el siguiente resultado establece que la suma de dos variables aleatorias independientes con distribución ji-cuadrada tiene distribución nuevamente ji-cuadrada con grados de libertad la suma de los grados de libertad de los sumandos.

Proposición. Si $X \sim \chi^2(n)$ y $Y \sim \chi^2(m)$ son dos variables aleatorias independientes, entonces $X + Y$ tiene distribución $\chi^2(n+m)$.

En particular, si $X_1, \ldots, X_n$ son independientes con distribución $\mathrm{N}(0,1)$, entonces $X_1^2 + \cdots + X_n^2$ tiene distribución $\chi^2(n)$.

DISTRIBUCIÓN t. Decimos que la variable aleatoria continua X tiene una distribución t con n grados de libertad si su función de densidad está dada por

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)}\,(1 + x^2/n)^{-(n+1)/2}, \qquad \text{para } -\infty < x < \infty.$$

En tal caso se escribe $X \sim t(n)$. La gráfica de esta función de densidad aparece en la Figura 1.31. Es posible demostrar que $E(X) = 0$, y $\mathrm{Var}(X) = n/(n-2)$ para $n > 2$. La distribución t se puede encontrar en los siguientes contextos.

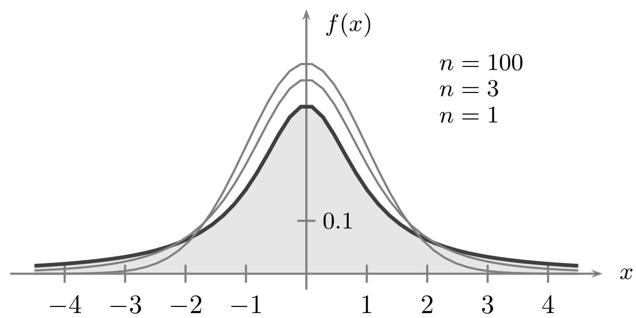


Figura 1.31:

Proposición. Si $X \sim \mathrm{N}(0,1)$ y $Y \sim \chi^2(n)$ son independientes, entonces

$$\frac{X}{\sqrt{Y/n}} \sim t(n).$$

Proposición. Sean $X_1, \ldots, X_n$ variables aleatorias independientes cada una de ellas con distribución $\mathrm{N}(\mu, \sigma^2)$. Entonces

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

en donde $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, y $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$.

Con esto terminamos con una revisión elemental de algunas distribuciones de probabilidad continuas. Recuerde el lector que existen muchas mas distribuciones de este tipo. La siguiente tabla contiene un resumen de las distribuciones continuas mencionadas en esta sección.

Distribución uniforme

$f(x) = 1/(b-a)$
para $x \in (a,b)$.
Parámetros: $a < b$.
Media: $(a+b)/2$.
Varianza: $(b-a)^2/12$.

Distribución exponencial

$f(x) = \lambda\, e^{-\lambda x}$
para $x > 0$.
Parámetro: $\lambda > 0$.
Media: $1/\lambda$.
Varianza: $1/\lambda^2$.

Distribución gama

$f(x) = \frac{(\lambda x)^{n-1}}{\Gamma(n)}\, \lambda\, e^{-\lambda x}$
para $x > 0$.
Parámetros: $n > 0, \lambda > 0$.
Media: n/λ.
Varianza: n/λ^2.

Distribución beta

$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$
para $x \in (0,1)$.
Parámetros: $a > 0, b > 0$.
Media: $a/(a+b)$.
Varianza: $ab/((a+b+1)(a+b)^2)$.

Distribución normal

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}$
para $x \in \mathbb{R}$.
Parámetros: $\mu \in \mathbb{R}, \sigma > 0$.
Media: μ.
Varianza: σ^2.

Distribución χ^2

$f(x) = \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2}$
para $x > 0$.
Parámetro: $n > 0$.
Media: n.
Varianza: $2n$.

Distribución t

$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)}\,(1+x^2/n)^{-(n+1)/2}$
para $x \in \mathbb{R}$.
Parámetro: $n > 0$.
Media: 0.
Varianza: $n/(n-2)$, para $n > 2$.

1.9. Vectores Aleatorios

Esta sección contiene una breve introducción al tema de *variables aleatorias multidimensionales* o también llamadas *vectores aleatorios*. Para hacer la escritura corta se consideran únicamente vectores aleatorios de dimensión dos, aunque todas las definiciones y resultados que se mencionan pueden extenderse fácilmente, en la mayoría de los casos, para vectores de dimensión superior.

Vector aleatorio. Un *vector aleatorio* de dimensión dos es un vector de la forma (X, Y) en donde cada coordenada es una variable aleatoria. De manera análoga se pueden tener vectores aleatorios multidimensionales $(X_1, \ldots, X_n)$. Nuevamente diremos que un vector aleatorio es *discreto*, o *continuo*, si las todas las variables aleatorias que lo conforman lo son. Por simplicidad consideraremos únicamente vectores aleatorios cuyas coordenadas son variables aleatorias todas discretas, o continuas, pero no mezclas de ellas. Un vector aleatorio (X, Y) puede considerarse como una función de Ω en $\mathbb{R}^2$ como se muestra en la Figura 1.32.

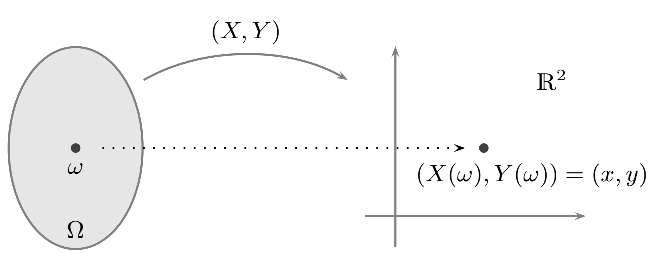


Figura 1.32:

Es decir, el vector (X, Y) evaluado en ω es $(X, Y)(\omega) = (X(\omega), Y(\omega))$ con posible valor (x, y). Nuevamente observe que el vector con letras mayúsculas (X, Y) es el vector aleatorio, mientras que el vector con letras minúsculas (x, y) es un punto en el plano. Estudiaremos a continuación algunas funciones asociadas a vectores aleatorios. Estos conceptos son completamente análogos al caso unidimensional estudiado antes.

Función de probabilidad conjunta. Vamos a estudiar primero el caso más sencillo que es el discreto. Consideremos un vector aleatorio discreto (X, Y) tal que la variable X toma los valores en el conjunto $\{x_1, x_2, \ldots\}$, y la variable Y toma los valores en $\{y_1, y_2, \ldots\}$. La función de densidad del vector (X, Y) que se denota por $f(x, y)$, se encuentra definida sobre $\mathbb{R}^2$ y toma valores en el intervalo $[0, 1]$ de acuerdo a la siguiente definición

$$f(x,y) = \begin{cases} P(X = x, Y = y) & \text{si } (x,y) \in \{x_1, x_2, \ldots\} \times \{y_1, y_2, \ldots\}, \\ 0 & \text{otro caso.} \end{cases}$$

Es decir, el número $f(x, y)$ es la probabilidad de que la variable X tome el valor x y al mismo tiempo la variable Y tome el valor y. Tal función se llama también *función de densidad conjunta* de las variables X y Y, y para enfatizar este hecho a veces se escribe $f_{X,Y}(x, y)$, pero en general omitiremos los subíndices para hacer la notación más corta. Toda función $f(x, y)$ de la forma anterior cumple las siguientes dos propiedades, e inversamente, toda función que las cumpla se llama función de densidad conjunta.

a) $f(x, y) \geq 0$.

b) $\displaystyle\sum_{x,y} f(x, y) = 1$.

Ejemplo. Considere el vector aleatorio discreto (X, Y) con función de densidad dada por la siguiente tabla.

x/y	0	1
-1	0.3	0.1
1	0.4	0.2

De este arreglo se entiende que la variable X toma valores en el conjunto $\{-1, 1\}$, mientras que Y toma valores en $\{0, 1\}$. Además las probabilidades conjuntas están dadas por las entradas de la tabla, por ejemplo, $P(X=-1,Y=0)=0.3$, esto es, la probabilidad de que X tome el valor -1 y al mismo tiempo Y tome el valor 0 es 0.3. La misma información puede escribirse de la siguiente manera.

$$f(x,y)=P(X=x,Y=y)=\begin{cases} 0.3 & \text{si } x=-1,\ y=0,\\ 0.1 & \text{si } x=-1,\ y=1,\\ 0.4 & \text{si } x=1,\ y=0,\\ 0.2 & \text{si } x=1,\ y=1,\\ 0 & \text{otro caso.}\end{cases}$$

Como todos estos valores son probabilidades, naturalmente son no negativos, y todos ellos suman uno. ∎

Función de densidad conjunta. Veamos ahora la situación en el caso continuo. Se dice que la función integrable y no negativa $f(x,y):\mathbb{R}^2\to[0,\infty)$ es la *función de densidad* del vector aleatorio continuo (X,Y) si para todo par (x,y) en $\mathbb{R}^2$ se cumple la igualdad

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

Toda función de densidad $f(x,y)$ de estas características satisface las siguientes dos propiedades que son análogas al caso discreto.

a) $f(x,y)\geq 0$.

b) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy=1$.

Recíprocamente decimos que una función $f(x,y):\mathbb{R}^2\to[0,\infty)$ es una *función de densidad conjunta* si cumple con las dos condiciones arriba mencionadas.

Ejemplo. Esta es una de las funciones de densidad conjunta más sencillas. Sean $a<b$, $c<d$, y defina la función

$$f(x,y)=\begin{cases} \dfrac{1}{(b-a)(d-c)} & \text{si } a<x<b,\ c<y<d,\\ 0 & \text{otro caso,}\end{cases}$$

cuya gráfica aparece en la Figura 1.33. Se trata de una función constante en el rectángulo $(a,b)\times(c,d)$. Esta función es de densidad pues es no negativa e integra uno sobre $\mathbb{R}^2$. ▪

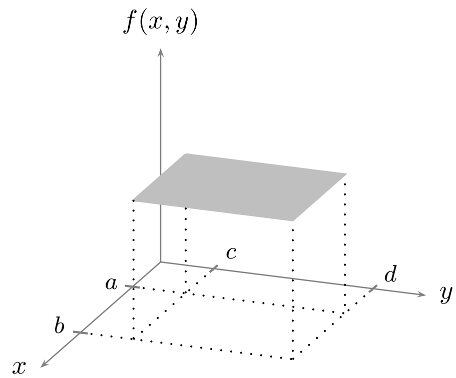


Figura 1.33:

Ejercicio. Compruebe que la siguiente función es de densidad.

$$f(x,y)=\begin{cases} x+y & \text{si } 0<x,y<1,\\ 0 & \text{otro caso.}\end{cases}$$

▪

Ejercicio. Encuentre la constante c para que la siguiente función sea de densidad.

$$f(x,y)=\begin{cases} cxy & \text{si } 0<x<y<1,\\ 0 & \text{otro caso.}\end{cases}$$

▪

Función de distribución conjunta. Además de la función de densidad, existe la función de distribución para un vector (X,Y), sea éste discreto o continuo, y su definición es muy semejante al caso unidimensional. La *función de distribución* del vector (X,Y), denotada por $F(x,y):\mathbb{R}^2\to[0,1]$, se define para cualquier par de números reales (x,y) como sigue

$$F(x,y)=P(X\leq x,Y\leq y).$$

La pequeña coma que aparece en el lado derecho de esta igualdad significa la intersección de los eventos $(X\leq x)$ y $(Y\leq y)$, es decir, el número $F(x,y)$ es la probabilidad del evento $(X\leq x)\cap(Y\leq y)$. Más precisamente, esta función debe escribirse como $F_{X,Y}(x,y)$, pero omitiremos los subíndices para mantener la notación simple. Siempre asociaremos la variable X con el valor x, y la variable Y con el valor y. A esta función se le conoce también con el nombre de *función de acumulación* de probabilidad del vector (X,Y), y también se dice que es la *función de distribución conjunta* de las variables X y Y.

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

1. $\displaystyle\lim_{x,y\to\infty}F(x,y)=1$.
2. $\displaystyle\lim_{x\to-\infty}F(x,y)=0$. Análogamente cuando es la variable y quien 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.
5. Para cualesquiera números $a<b$, y $c<d$, se cumple la desigualdad
$$F(b,d)-F(a,d)-F(b,c)+F(a,c)\geq 0.$$

Observe que las primeras cuatro propiedades son análogas al caso unidimensional, y puede comprobarse geométricamente que la quinta propiedad corresponde a la probabilidad del evento $(a<X\leq b)\cap(c<Y\leq d)$. Recíprocamente decimos que una función $F(x,y):\mathbb{R}^2\to[0,1]$ es una *función de distribución bivariada* si satisface las anteriores cinco propiedades.

El concepto de función de distribución puede extenderse al caso de vectores multidimensionales de la siguiente forma: La función de distribución del vector aleatorio $(X_1,\ldots,X_n)$ es la función $F(x_1,\ldots,x_n):\mathbb{R}^n\to[0,1]$ dada por

$$F(x_1,\ldots,x_n)=P(X_1\leq x_1,\ldots,X_n\leq x_n).$$

a) ¿Cómo se encuentra $F(x,y)$ a partir de $f(x,y)$? Conociendo la función de densidad $f(x,y)$, es posible encontrar al función de distribución $F(x,y)$ simplemente integrando en el caso continuo o sumando en el caso discreto. Para el caso continuo tenemos

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, dv\, du.$$

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

b) ¿Cómo se encuentra $f(x,y)$ a partir de $F(x,y)$? En el caso continuo sabemos que $f(x,y)$ y $F(x,y)$ guardan la relación

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, dv\, du,$$

por el teorema fundamental del cálculo tenemos que

$$f(x,y) = \frac{\partial^2}{\partial x\, \partial y} F(x,y).$$

Función de densidad marginal. Sea $f(x,y)$ la función de densidad del vector aleatorio continuo (X,Y). Se define la *función de densidad marginal* de la variable X como sigue

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\, dy,$$

es decir, simplemente se integra respecto de la variable y. La correspondiente función de densidad marginal de la variable Y se obtiene integrando ahora respecto de la variable x, es decir,

$$f(y) = \int_{-\infty}^{\infty} f(x,y)\, dx.$$

La correspondiente definición para vectores discretos involucra una suma en lugar de la integral, por ejemplo,

$$f(x) = \sum_{y} f(x,y),$$

de manera análoga se define la densidad marginal $f(y)$. Es sencillo verificar que estas densidades marginales, tanto en el caso discreto como en el continuo, son efectivamente funciones de densidad univariadas. Un poco más generalmente, la función de densidad marginal de la variable X_1 a partir de la función de densidad del vector $(X_1, \ldots, X_n)$ es, en el caso continuo,

$$f(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_2 \cdots dx_n.$$

De manera análoga se obtiene la función marginal de cualquiera de las variables que componen el vector multidimensional. Y también de manera análoga se pueden calcular estas densidades marginales para vectores discretos de cualquier dimensión.

Función de distribución marginal. Sea (X, Y) un vector aleatorio, continuo o discreto, con función de distribución $F(x, y)$. La *función de distribución marginal* de la variable X se define como la función

$$F(x) = \lim_{y \to \infty} F(x, y),$$

y la correspondiente función de distribución marginal de la variable Y es la función

$$F(y) = \lim_{x \to \infty} F(x, y).$$

No es difícil comprobar que estas funciones de distribución marginales son efectivamente funciones de distribución univariadas.

Todo lo mencionado en estas últimas secciones tiene como objetivo poder enunciar con precisión el concepto de independencia entre variables aleatorias. Veremos a continuación este importante concepto que usaremos con regularidad en la segunda parte del curso.

Independencia de variables aleatorias. Sean $X_1, \ldots, X_n$ variables aleatorias con distribución conjunta $F(x_1, \ldots, x_n)$, y con respectivas distribuciones marginales $F(x_1), \ldots, F(x_n)$. Se dice que estas variables son *independientes* si para cualesquiera valores $x_1, \ldots, x_n$ se cumple la igualdad

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n).$$

Alternativamente puede definirse la independencia en términos de la función de densidad como sigue

$$f(x_1, \ldots, x_n) = f(x_1) \cdots f(x_n).$$

Nuevamente esta igualdad debe verificarse para cualesquiera valores $x_1, \ldots, x_n$. Las dos condiciones anteriores son equivalentes. En el caso particular cuando las variables aleatorias son discretas, la condición de independencia se escribe

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n).$$

Adicionalmente, se dice que un conjunto infinito de variables aleatorias son independientes si cualquier subconjunto finito de ellas lo es. Este es el sentido en el que la sucesión infinita de variables aleatorias debe entenderse en el enunciado del teorema central del límite.

Parte 2

ESTADÍSTICA

Originalmente cualquier colección de datos acerca de la población o economía de un estado era del interés exclusivo del estado, es por ello que a las primeras colecciones de datos se les llamara *estadísticas*, como una derivación de la palabra *estado*. Los primeros registros de una recolección sistemática de datos de una población surgen en las ciudades italianas de Venecia y Florencia durante el Renacimiento. Para mediados del siglo XVI, varios gobiernos europeos, a través de las parroquias, llevaban registros de nacimientos, matrimonios y decesos. De particular importancia en aquellos años eran los registros de decesos pues las tasas de mortalidad eran altas debido a las recurrentes epidemias y plagas que diezmaban a la población. Esta recolección de datos por parte de los gobiernos siguió desarollándose durante los siglos XVII y XVIII, en donde los esfuerzos y recursos se orientaron principalmente a obtener censos.

A finales del siglo IXX, y teniendo a Francis Galton (1822-1911) como precursor, surgen las primeras ideas para inferir información de una población a partir de datos estadísticos. Los métodos y la teoría general se desarrollaron a partir de los inicios del siglo XX gracias a los trabajos de Karl Pearson (1857-1936), Ronald A. Fisher (1890-1962), Egon Sharpe Pearson (1895-1980) hijo de Karl Pearson, y Jerzy Neymann (1894-1981).

Hoy en día los métodos de la estadística son usados en muy diversas áreas del conocimiento humano, sean éstas científicas, sociales o económicas. Sus métodos y conclusiones son usados como respaldo científico para rechazar o aceptar afirmaciones, y para la toma de decisiones. Cotidianamente encontramos elementos de estadística descriptiva en los periódicos, en la radio, en la televisión y en internet.

2.1. Introducción

Supongamos que tenemos una *población* de interés, esto es, un conjunto arbitrario de personas, mediciones u objetos cualesquiera, y deseamos conocer cierta información de esta población. Debido a la imposibilidad o no conveniencia de tener información de todos y cada uno de los elementos de la población, tomamos un pequeño subconjunto de ellos, al cual llamaremos *muestra*. Con base en esta muestra trataremos de inferir la información de la población en su totalidad.

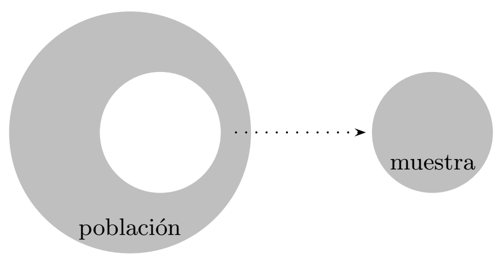


Figura 2.1:

Estadística descriptiva e inferencial. La estadística es la ciencia que se encarga de recolectar, organizar, resumir y analizar datos para después obtener conclusiones a partir de ellos. De manera general, la estadística puede ser dividida en dos grandes áreas: estadística inferencial y estadística descriptiva. La *estadística descriptiva* es una colección de métodos para la organización, resumen y presentación de datos. La *estadística inferencial* consiste de un conjunto de técnicas para obtener, con determinado grado de confianza, información de una población con base en la información de una muestra.

2.2. Variables y tipos de datos

Una *variable* es una característica de un elemento en una población en estudio. Por ejemplo, si la población consta de personas, los siguientes son ejemplos de variables que podrían ser de interés: edad, peso, sexo, estatura, etc. Veremos a continuación algunos criterios bajo los cuales las variables pueden ser clasificadas.

Las variables pueden ser *cuantitativas*, cuando se realiza una medición y el resultado es un número, o pueden ser *cualitativas*, cuando solamente registran una cualidad o atributo del objeto o persona en estudio. La edad, el peso y la estatura son ejemplos de variables cuantitativas en una población de personas, mientras que el sexo y el estado civil son variables cualitativas.

De acuerdo al número total de sus posibles valores, una variable cuantitativa puede ser clasificada como *discreta* cuando sólo puede tomar un número discreto (es decir, finito o numerable) de valores, o *continua* cuando puede tomar cualquier valor dentro de un intervalo (a, b) de la recta real.

De acuerdo con la posible relación que pudieran guardar los valores de una variable, se cuenta por lo menos con cuatro escalas de medición. Las variables cualitativas pueden ser clasificadas de acuerdo a dos escalas: escala nominal o escala ordinal. Mientras que las variables cuantitativas pueden clasificarse por: escala de intervalo o escala de razón.

Escala nominal. Una variable se llama *nominal* cuando sus posibles valores no tienen alguna relación de orden o magnitud entre ellos. Básicamente los valores de este tipo de variables son etiquetas sin un orden entre ellos. Por ejemplo, si estamos estudiando una población humana, a la variable sexo podemos asignarle dos posibles valores: F para femenino, y M para masculino. Los símbolos F y M son etiquetas arbitrarias, y no existe un orden en ellas ni podemos realizar operaciones aritméticas. La religión o la nacionalidad son también ejemplos de variables nominales.

Escala ordinal. En esta escala los valores de la variable tienen un orden pero no se pueden hacer operaciones aritméticas entre estos valores pues no hay noción de distancia entre ellos. Por ejemplo, para calificar las características de un objeto podemos suponer los siguientes valores: 0=Pésimo, 1=malo, 2=Regular, 3=Bueno, 4=Excelente. En este caso la escala de medición es ordinal pues existe un orden entre sus valores, pero no se puede decir, por ejemplo, que dos valores regulares hacen un valor excelente.

Escala de intervalo. En este tipo de escala existe un orden entre los valores de la variable y existe además una noción de distancia aunque no se pueden realizar operaciones. No existe el valor natural cero para esta tipo de escala. Por ejemplo, suponga que los valores de una cierta variable están dados por los días del mes. Entre el día 10 y el día 20 hay una distancia de diez días, pero no se puede decir que el día 20 es dos veces el día 10. La temperatura es otro ejemplo de este tipo de variable, el posible valor cero depende de la escala que se use para medir la temperatura (Celsius, Kelvin, Fahrenheit).

Escala de razón. En una escala de razón la magnitud tiene un sentido físico y existe el cero absoluto. Por ejemplo, la variable edad en años estudiada en una población humana.

La clasificación de un variable particular en alguna de estas categorías puede no ser clara pues tal decisión puede depender del tratamiento que una persona haga de tal variable.

Ejercicio. Determine si cada una de las siguientes variables es cualitativa o cuantitativa. Si es cualitativa, diga si su escala de medición es nominal u ordinal. Si es cuantitativa, diga si es discreta o continua, y diga su escala de medición.

a) Apreciación o punto de vista acerca de una idea o propuesta usando la escala: De Acuerdo, En desacuerdo, Indiferente.

b) Evaluación de un curso de acuerdo a las siguientes categorías: Acreditado, No Acreditado, Ausente.

c) Número de hijos en una familia.

d) Salario en unidades monetarias.

e) Nivel de salario en alguna de las categorías: A,B,C,D,E,F.

f) Longitud de un tornillo.

g) Preferencia sexual.

h) Lugar de nacimiento de una persona.

i) Peso de un bebé recién nacido. ▪

2.3. Estadística descriptiva

Supongamos que tenemos un conjunto de datos numéricos $x_1, \ldots, x_n$, que representan mediciones de alguna variable de interés en un experimento aleatorio. Para conocer algunas características globales de esta variable se pueden calcular ciertas medidas de tendencia central como la media, moda y mediana; y también otras medidas llamadas de dispersión como la varianza, la desviación estándar y el rango. Definiremos estos conceptos a continuación.

Medidas de tendencia central. La *media* de los datos numéricos $x_1, \ldots, x_n$, denotada por $\bar{x}$, es simplemente el promedio $(x_1 + \cdots + x_n)/n$. Por otro lado, la *moda* es el valor que aparece con mayor frecuencia. Si ningún valor se repite, se dice que no hay moda, o que todos los valores son moda. Si existe un único valor con mayor número de repeticiones, entonces a ese valor se le llama la moda, y el conjunto de datos se dice que es unimodal. Pueden existir, si embargo, dos o mas valores que sean los que mayor número de veces se repiten, en tal caso la moda consta de esos valores, y se dice que el conjunto de datos es bimodal. Si hay mas de dos valores con mayor número de repeticiones, se dice que el conjunto de datos es multimodal. Para calcular la mediana procedemos como sigue. Se ordena la muestra $x_1, \ldots, x_n$ de menor a mayor incluyendo repeticiones, y se obtiene la muestra ordenada $x_{(1)}, \ldots, x_{(n)}$, en donde $x_{(1)}$ denota el dato más pequeño y $x_{(n)}$ es el dato más grande. La *mediana*, denotada por $\tilde{x}$, se define como sigue

$$\tilde{x} = \begin{cases} \frac{1}{2}[x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}] & \text{si } n \text{ es par,} \\ x_{(\frac{n+1}{2})} & \text{si } n \text{ es impar.} \end{cases}$$

De este modo, cuando tenemos un número impar de datos, la mediana es el dato ordenado que se encuentra justo a la mitad. Y cuando tenemos un número par de datos, la mediana se calcula promediando los dos datos ordenados que están enmedio.

Ejercicio. Calcule la media, moda y mediana de cada uno de los siguientes conjuntos de datos.

a) $-5, 2, -3, -2, 4, 4, 2, 5$.

b) $3.5, 2, 7.5, 3.5, 1.5, 4.5, 7.5$.

c) $1, 2, 3, 4, 5, 6, 7, 8, 9, 10$. ▪

Medidas de dispersión. Definiremos ahora algunas medidas que ayudan a cuantificar la dispersión que puede presentar un conjunto de datos numéricos $x_1, \ldots, x_n$. La *varianza*, denotada por s^2, se define como sigue

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

en donde $\bar{x}$ es la media muestral definida antes. La *desviación estándar* es la raíz cuadrada positiva de s^2, y se le denota naturalmente por s. El *rango* es el dato más grande menos el dato más pequeño, es decir, $x_{(n)} - x_{(1)}$.

Ejercicio. Calcule la varianza, la desviación estándar y el rango de los conjuntos de datos del ejercicio anterior. ▪

2.4. Muestras aleatorias y estadísticas

En lo que los problemas estadísticos que consideraremos no estudiaremos muestras particulares de datos numéricos $x_1, \ldots, x_n$, sino conjuntos de variables aleatorias $X_1, \ldots, X_n$ que pueden, en particular, tomar el conjunto de datos mencionado. Definiremos a continuación este concepto junto con el de estadística.

Definición. Una *muestra aleatoria* (escribimos simplemente m.a.) es una colección de variables aleatorias $X_1, \ldots, X_n$ que son independientes e idénticamente distribuidas.

De este modo, cuando se diga, por ejemplo, que una muestra aleatoria es tomada de una población normal con media μ y varianza σ^2, ello significa que las variables aleatorias que forman la m.a. son independientes entre sí, y todas ellas tienen la misma distribución normal con los mismos parámetros. Una muestra aleatoria constituye el elemento básico para llevar a cabo inferencias estadísticas.

Definición. Una *estadística* es una función cualquiera de una muestra aleatoria $X_1, \ldots, X_n$, y por lo tanto es también una variable aleatoria.

Veremos a continuación dos ejemplos de estadísticas que serán usados con frecuencia más adelante. Considere una muestra aleatoria $X_1, \ldots, X_n$. La función $\bar{X}$ definida como sigue

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

es una estadística, y se le conoce con el nombre de *media muestral*. El otro ejemplo es el de la *varianza muestral*, que se define de la forma siguiente

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

2.5. Estimación puntual

Sea X una variable aleatoria de interés en un experimento aleatorio, y supongamos que X tiene una distribución de probabilidad con función de densidad $f(x;\theta)$, en donde θ es el parámetro o el conjunto de parámetros de la distribución. En esta nueva notación se hace énfasis en que la distribución depende de un parámetro denotado genéricamente por la letra griega θ, y el cual consideraremos desconocido. Por ejemplo, si la distribución es $\exp(\lambda)$, entonces θ representa el parámetro λ, si la distribución es $\mathrm{N}(\mu,\sigma^2)$, entonces θ representa el vector de parámetros (μ,σ^2). El problema de estimación puntual consiste en encontrar un número, con base en las observaciones realizadas de la variable aleatoria, que sirva como estimación del parámetro desconocido θ. Ilustraremos el problema con un par de ejemplos.

Ejemplo. Se desea conocer el estado de la calidad de un conjunto 1000 artículos. Dada la imposibilidad de someter a prueba a todos ellos, se escogen 20 artículos al azar obteniéndose los siguientes resultados: 0,1,1,0,1,1,0,1,0,1,1,1,0,1,1,1,1,1,1,0, en donde cero indica que el artículo no pasó el control de calidad y uno indica que el artículo pasó el control de calidad. Suponga que X es la variable que indica si un artículo escogido al azar pasa o no pasa el control de calidad. Entonces X tiene una distribución $\mathrm{Ber}(p)$, en donde p es desconocido. ¿Cómo podría estimarse el verdadero valor de p con base en los datos de la muestra? ▪

Ejemplo. El tiempo en minutos que un conjunto de 10 empleados escogidos al azar invierte en trasladarse de la casa al lugar de trabajo es: 30,70,65,10,25,15,5,50,20,15. Suponga que tal variable puede modelarse mediante la distribución $\exp(\lambda)$. ¿Cómo podría estimarse el valor de λ con base en las observaciones realizadas? ▪

Definición. Un *estimador puntual* para el parámetro θ es una función de una muestra aleatoria $X_1,\ldots,X_n$ que se usa para estimar θ.

A un estimador del parámetro θ se le denota regularmente por $\hat{\theta}$ (se lee "teta circunflejo"). Observe que un estimador puntual es una estadística y puede escribirse como $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$. Veremos a continuación dos métodos para encontrar estimadores puntuales.

Método de máxima verosimilitud. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con función de densidad $f(x;\theta)$. Esto significa que todas las variables de la muestra aleatoria tienen función de densidad $f(x)$ que depende de un parámetro desconocido θ.

Definición. La función de verosimilitud de una muestra aleatoria $X_1, \ldots, X_n$, denotada por $L(\theta)$, se define como la función de densidad conjunta

$$L(\theta) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta).$$

La letra L proviene del término en inglés *likelihood*, que tradicionalmente se ha traducido como *verosimilitud*, aunque tal vez el término *credibilidad* sea más acertado. El método de máxima verosimilitud consiste en obtener el valor de θ que maximice la función de verosimilitud $L(\theta)$, la idea intuitiva es interesante: se debe encontrar el valor de θ de tal forma que los datos observados tengan máxima probabilidad de ocurrir. El valor de θ en donde se alcanza el máximo se llama estimador de *máxima verosimilitud*, o estimador *máximo verosímil*. Ilustraremos este método con algunos ejemplos.

Ejemplo. Encontraremos el estimador máximo verosímil para el parámetro λ de una distribución exponencial. La función de verosimilitud es

$$\begin{aligned} L(\lambda) &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ &= \lambda\, e^{-\lambda x_1} \cdots \lambda\, e^{-\lambda x_n} \\ &= \lambda^n\, e^{-\lambda n \bar{x}} \end{aligned}$$

Maximizar la función $L(\lambda)$ es equivalente a maximizar la función $\ln L(\lambda)$, pues la función logaritmo es continua y monótona creciente en su dominio de definición. Hacemos esto pues esta nueva función resulta más fácil de maximizar como veremos a continuación. Tenemos que $\ln L(\lambda) = n \ln \lambda - \lambda n \bar{x}$. Derivando respecto a λ e igualando a cero se llega a la ecuación $n/\lambda - n\bar{x} = 0$, de donde se obtiene $\lambda = 1/\bar{x}$. El estimador es entonces $\hat{\lambda} = 1/\bar{X}$. ▪

Ejemplo. Encontraremos los estimadores de máxima verosimilitud para los parámetros μ y σ^2 de una distribución normal. Por definición la función de verosimilitud es

$$\begin{aligned} L(\mu,\sigma^2) &= f_{X_1}(x_1)\cdots f_{X_n}(x_n) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x_1-\mu)^2/2\sigma^2}\cdots\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x_n-\mu)^2/2\sigma^2} \\ &= (\frac{1}{\sqrt{2\pi\sigma^2}})^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2}. \end{aligned}$$

Nuevamente el logaritmo de esta función es más sencillo de maximizar. Tenemos que

$$\ln L(\mu,\sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2.$$

Por lo tanto

$$\begin{aligned} \frac{\partial}{\partial\mu}\ln L(\mu,\sigma^2) &= \frac{1}{\sigma^2}\sum_{i=1}^n(x_i-\mu), \\ \frac{\partial}{\partial\sigma^2}\ln L(\mu,\sigma^2) &= -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^n(x_i-\mu)^2. \end{aligned}$$

Igualando a cero ambas derivadas encontramos un sistema de dos ecuaciones con dos variables:

$$\begin{aligned} \frac{1}{\sigma^2}\sum_{i=1}^n(x_i-\mu) &= 0, \\ -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^n(x_i-\mu)^2 &= 0. \end{aligned}$$

De estas ecuaciones se obtiene $\mu = \frac{1}{n}\sum_{i=1}^n x_i$, y $\sigma^2 = \frac{1}{n}\sum_{i=1}^n(x_i-\hat{\mu})^2$. Por lo tanto los estimadores para los parámetros μ y σ^2 de una distribución normal por el método de máxima verosimilitud son $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$, y $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n(X_i-\hat{\mu})^2$. ▪

Método de momentos. Sea nuevamente $f(x;\theta)$ la función de densidad de una variable aleatoria X que depende de un parámetro desconocido θ. Recordemos que el k-ésimo *momento poblacional* de X es el número $E(X^k)$, cuando esta esperanza existe. Ahora, dada una muestra aleatoria $X_1, \ldots, X_n$ de esta distribución, se define el k-ésimo *momento muestral* como $\frac{1}{n}\sum_{i=1}^{n} X_i^k$. El *método de momentos* para estimar el parámetro θ es muy sencillo, consiste en igualar los momentos muestrales con los correspondientes momentos poblacionales, y resolver esta ecuación o sistema de ecuaciones para el parámetro θ cuando ello sea posible. Veamos algunos ejemplos.

Ejemplo. Nuevamente estimaremos los parámetros μ y σ^2 de una distribución normal. Esta vez usaremos el método de momentos. Como necesitamos estimar dos parámetros usamos los dos primeros momentos. El primer y segundo momento poblacionales son $E(X) = \mu$ y $E(X^2) = \sigma^2 + \mu^2$, respectivamente. El primer y segundo momento muestrales son $\sum_{i=1}^{n} x_i$ y $\sum_{i=1}^{n} x_i^2$, respectivamente. La igualación respectiva produce el sistema de ecuaciones

$$\begin{aligned} \mu &= \frac{1}{n}\sum_{i=1}^{n} x_i, \\ \sigma^2 + \mu &= \frac{1}{n}\sum_{i=1}^{n} x_i^2. \end{aligned}$$

La primera ecuación es explícita mientras que la segunda ecuación se puede reescribir como sigue

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\frac{1}{n}\sum_{i=1}^{n} x_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

En este caso los estimadores por el método de momentos coinciden con los estimadores máximo verosímiles. ▪

Insesgamiento. Siendo un estimador $\hat{\theta}$ una variable aleatoria que se utiliza para estimar el parámetro θ, es interesante saber si el valor promedio de $\hat{\theta}$ es θ. Esta sería una buena propiedad para un estimador.

Definición. Un estimador $\hat{\theta}$ del parámetro θ es *insesgado* si $E(\hat{\theta}) = \theta$.

Si $\hat{\theta}$ no es insesgado, entonces se dice que es *sesgado*, y a la diferencia $E(\hat{\theta}) - \theta$ se le llama *sesgo*. De esta forma, un estimador puntual $\hat{\theta}$ es un estimador insesgado para el parámetro desconocido θ si, en promedio, el valor de $\hat{\theta}$ coincide con el valor desconocido de θ. Observe en los siguientes ejemplos la forma en la que se verifica la propiedad de insesgamiento aún cuando no se conozca el valor del parámetro θ.

Ejemplo. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con media desconocida μ. Comprobaremos que la media muestral $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ es un estimador insesgado para el parámetro μ. Observe que $\bar{X}$ es el estimador $\hat{\theta}$, y μ es el parámetro desconocido θ. Por la propiedad lineal de la esperanza,

$$E(\hat{\mu}) = E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu.$$

▪

Ejemplo. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con varianza desconocida σ^2. Recordemos que la varianza muestral es una estadística definida de la forma siguiente: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. En este caso el estimador es S^2 y el parámetro desconocido a estimar es σ^2. Esta estadística resulta ser un estimador insesgado para la varianza σ^2. Comprobar esta afirmación requiere de algunos cálculos que por cuestión de espacio omitimos, de modo que se deja al lector tal comprobabción, lo único que hay que hacer es desarrollar el cuadrado, usar la propiedad de linealidad de la esperanza y observar que

$$E(X_i X_j) = \begin{cases} \mu^2 & \text{si } i \neq j, \\ \sigma^2 + \mu^2 & \text{si } i = j. \end{cases}$$

▪

2.6. Estimación por intervalos

En algunos casos es preferible no dar un número como estimación de un parámetro sino un intervalo de posibles valores. En esta sección se estudia brevemente el tema de estimación de parámetros usando intervalos. En este tipo de estimación se busca un intervalo de tal forma que se pueda decir, con cierto grado de confiabilidad, que dicho intervalo contiene el verdadero valor del parámetro desconocido. A este tipo de intervalos se les llama *intervalos de confianza*.

Definición. Sea $\alpha \in (0,1)$. Un intervalo de confianza para un parámetro desconocido θ de una distribución de probabilidad es un intervalo aleatorio de la forma $(\hat{\theta}_1, \hat{\theta}_2)$, en donde $\hat{\theta}_1$ y $\hat{\theta}_2$ son estadísticas (funciones de una muestra aleatoria) tales que

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha. \tag{2.1}$$

A las estadísticas $\hat{\theta}_1$ y $\hat{\theta}_2$ se les conoce como *límites inferior y superior*, respectivamente, del intervalo de confianza. Al número $1-\alpha$ se le conoce como *grado* o *coeficiente de confianza.* En general, se toma el valor de α cercano a 0 de tal forma que el grado de confianza, $1-\alpha$, es cercano a 1. En la práctica es común tomar α =0.05, de modo que el grado de confianza es $1-\alpha$ =0.95 . Decimos entonces que el grado de confianza es del 95 %. Observe que las estadísticas $\hat{\theta}_1$ y $\hat{\theta}_2$ dependen de una muestra aleatoria $X_1, \ldots, X_n$, de modo que al tomar estas variables aleatorias distintos valores se generan distintos intervalos de confianza. Esto se ilustra en la Figura 2.2.

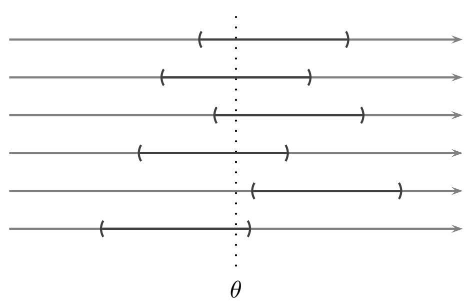


Figura 2.2:

Observe que no es correcto decir "la probabilidad de que θ pertenezca al intervalo $(\hat{\theta}_1, \hat{\theta}_2)$ es $1-\alpha$", pues el parámetro no es el elemento aleatorio. En cambio, se dice "la probabilidad de que el intervalo $(\hat{\theta}_1, \hat{\theta}_2)$ contenga el valor de θ es $1-\alpha$". De esta forma se entiende que θ es constante, aunque desconocido, y el intervalo es el que cambia dependiendo de la muestra. Naturalmente el problema es encontrar $\hat{\theta}_1$ y $\hat{\theta}_2$ de tal forma que la igualdad (2.1) se cumpla. A continuación mostraremos la forma de resolver este problema en algunos casos particulares.

Intervalo para la media de una distribución normal con varianza conocida. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal con media desconocida μ y varianza conocida σ^2. Encontraremos un intervalo de confianza para el parámetro μ. Como cada una de las variables de la muestra tiene distribución $\mathrm{N}(\mu, \sigma^2)$, la media muestral $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ tiene distribución $\mathrm{N}(\mu, \sigma^2/n)$. De modo que, estandarizando,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1).$$

Para cualquier valor de $\alpha \in (0,1)$ podemos encontrar un valor $z_{\alpha/2}$ en tablas de probabilidad normal estándar, véase la Figura 2.3, tal que

$$P(\,-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\,) = 1 - \alpha.$$

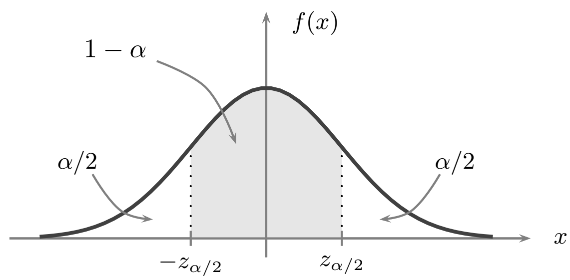


Figura 2.3:

Despejando la constante desconocida μ se obtiene el siguiente resultado.

Proposición. Un intervalo de confianza para la media μ de una distribución normal con varianza conocida σ^2 está dado por la siguiente expresión

$$P(\,\bar{X} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\,) = 1 - \alpha. \tag{2.2}$$

De esta forma, el intervalo $(\bar{X} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}})$ es un intervalo de confianza para el parámetro desconocido μ, pues contiene a dicho parámetro con probabilidad $1 - \alpha$. Observe que todas las expresiones que aparecen en este intervalo son conocidas. Ilustraremos la aplicación de esta fórmula mediante un ejemplo.

Ejemplo. Suponga que la vida promedio útil, medida en horas, de focos de 100 watts producidos por cierta compañía, puede ser modelada mediante una variable aleatoria con distribución normal de media μ y varianza σ^2. Suponga que la desviación estándar σ es conocida y es igual a 30 horas. El objetivo es encontrar un intervalo de confianza para la vida promedio útil μ de los focos producidos por esta compañía. Para ello se toma una muestra de 20 focos y mediante pruebas de laboratorio se determina la vida útil de cada uno de ellos. Los resultados $x_1, \ldots, x_{20}$ arrojan una media muestral $\bar{x}$ de 1050 horas. Si consideramos un nivel de confianza del 95 %, es decir, α =0.05, de la tabla de probabilidad normal se encuentra que $z_{\alpha/2} = z_{0.025}$ =1.96, y entonces puede ahora calcularse el intervalo

$$\begin{aligned}
(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) &= (1050 - 1.96\,\frac{30}{\sqrt{20}},\ 1050 + 1.96 \times \frac{30}{\sqrt{20}}) \\
&= (1050 - 13.148,\ 1050 + 13.148) \\
&= (1036.852,\ 1063.148).
\end{aligned}$$

De esta forma, con una confianza del 95 %, la vida promedio útil de este tipo de focos es de 1050 ± 13.148 horas. ▪

Ejercicio. Compruebe que la longitud del intervalo aleatorio que aparece en (2.2) es $2z_{\alpha/2}\sigma/\sqrt{n}$. Observe que la longitud del intervalo decrece conforme el tamaño de la muestra crece. Además, si la confianza requerida crece, es decir, si $1 - \alpha$ aumenta, entonces $z_{\alpha/2}$ crece (véase la Figura 2.3), y por lo tanto la longitud del intervalo también crece. ▪

Ejercicio. Encuentre un intervalo de confianza al 90 % para la media de una población normal con $\sigma = 5$ cuando se ha tomado una muestra de tamaño 25 cuya media muestral es 60. ▪

Intervalo para la media de una distribución normal con varianza desconocida. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución normal con media desconocida μ y varianza desconocida σ^2. El resultado teórico fundamental en la siguiente derivación es que la variable aleatoria

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

tiene una distribución t con $n - 1$ grados de libertad. Observe que esta es la distribución exacta de la variable T, sin importar el tamaño de la muestra y sobre todo, sin suponer que la varianza de la muestra es conocida. A partir de lo anterior podemos construir un intervalo de confianza para el parámetro desconocido μ de forma análoga al caso normal mencionado antes. Para cualquier valor de $\alpha \in (0,1)$ podemos encontrar un valor $t_{\alpha/2}$ en tablas de probabilidad de la distribución t de $n-1$ grados de libertad (véase la Figura 2.4) tal que

$$P(\,-t_{\alpha/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2}\,) = 1-\alpha.$$

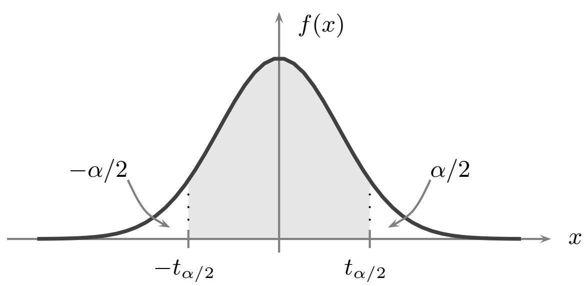


Figura 2.4:

Despejando la constante desconocida μ de la ecuación anterior se obtiene el siguiente resultado.

Proposición. Un intervalo de confianza para la media μ de una distribución normal está dado por la siguiente expresión

$$P(\,\bar{X} - t_{\alpha/2}\,\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\,\frac{S}{\sqrt{n}}\,) = 1-\alpha. \tag{2.3}$$

De este modo, el intervalo $(\,\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\,)$ es un intervalo de confianza para la media μ de una población normal sin suponer la varianza conocida. No es explícito en la notación pero el valor $t_{\alpha/2}$ corresponde a la distribución t con $n-1$ grados de libertad, a veces se escribe $t_{\alpha/2,n-1}$.

Intervalo aproximado para la media de una distribución cualquiera. Sea $X_1,\ldots,X_n$ una muestra aleatoria de una distribución cualquiera con media desconocida μ. Supongamos que el tamaño n de la muestra es grande, i.e. $n \geq 30$.

Entonces, por el teorema central del límite, la variable aleatoria

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

tiene una distribución aproximada normal estándar. Se puede encontrar un intervalo aproximado de confianza para el parámetro desconocido μ siguiendo el procedimiento antes realizado. Para cualquier valor de $\alpha \in (0,1)$ podemos encontrar un valor $z_{\alpha/2}$ en tablas de probabilidad normal estándar tal que

$$P(\, -z_{\alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < z_{\alpha/2}\,) = 1 - \alpha.$$

Despejando nuevamente la constante desconocida μse obtiene

$$P(\, \bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\,) = 1 - \alpha.$$

De esta forma, el intervalo $(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}})$ es un intervalo de confianza aproximado para el parámetro desconocido μ pues contiene a dicho parámetro con probabilidad $1 - \alpha$. Observe nuevamente que todas las expresiones que aparecen en este intervalo son conocidas.

A manera de resumen se tiene la siguiente tabla.

Hipótesis	Intervalo para la media μ
Distribución normal con varianza σ^2 conocida	$P(\, \bar{X} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\,) = 1 - \alpha.$
Distribución normal con varianza σ^2 desconocida	$P(\, \bar{X} - t_{\alpha/2,n-1}\,\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\,\frac{S}{\sqrt{n}}\,) = 1 - \alpha.$
Cualquier distribución Muestra grande, $n \geq 30$	Intervalo aproximado: $P(\, \bar{X} - z_{\alpha/2}\,\frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\,\frac{S}{\sqrt{n}}\,) = 1 - \alpha.$

2.7. Pruebas de hipótesis

Ilustraremos la idea de una prueba de hipótesis con un ejemplo. Considere una situación en la que se efectúa sólo uno de los siguientes dos experimentos aleatorios: En uno de ellos se lanza un dado, en el otro se lanzan cinco monedas y se cuenta el número de cruces totales que se obtienen, suponiendo que los lados de cada moneda se denominan cara o cruz. Suponga que únicamente conocemos el resultado de uno u otro experimento, y se nos pide determinar cuál de los dos experimentos se realizó con base en el número reportado. Tenemos entonces una situación de dos hipótesis

$$H_0 : \text{Dado} \quad vs \quad H_1 : \text{Monedas}.$$

Como información se tiene un número dentro del conjunto $\Omega = \{0, 1, 2, 3, 4, 5, 6\}$. Veamos la situación. Si el número reportado ω es 0, entonces seguro se realizó el experimento de las monedas. Si $\omega = 6$, entonces con seguridad el dado fue lanzado. ¿Qué decisión tomar para cualquier otro valor de ω? En la siguiente tabla se muestran las probabilidades de los posibles números reportados bajo cada uno de los dos experimentos.

ω	0	1	2	3	4	5	6
H_0 : Dado	0	$1/6^*$	$1/6$	$1/6$	$1/6^*$	$1/6^*$	$1/6^*$
H_1 : Monedas	$1/32^*$	$5/32$	$10/32^*$	$10/32^*$	$5/32$	$1/32$	0

La estrategia natural es decidir por el experimento que tenga mayor probabilidad para cada valor de ω. Estas probabilidades mayores se encuentran indicadas con un asterisco en la tabla anterior. De esta forma se llega a la siguiente regla de decisión:

Si $\omega \in \mathscr{C} = \{0, 2, 3\}$, entonces se rechaza H_0, en caso contrario se acepta H_0.

Por razones naturales al conjunto $\mathscr{C}$ se le llama *región del rechazo* de la hipótesis H_0. La regla de decisión anterior es razonable, sin embargo no está libre de errores. Si $\omega = 2$, se decide por las monedas, pero el resultado bien pudo provenir del dado. Igualmente, si $\omega = 1$, se decide por el dado pero es factible que el resultado haya sido obtenido por las monedas. Se tiene entonces las siguientes definiciones:

Error tipo I: Rechazar H_0 cuando H_0 es verdadera.

Error tipo II: No rechazar H_0 cuando H_0 es falsa.

Las probabilidades de estos errores pueden ser calculadas del siguiente modo.

$$\begin{aligned} P(\text{“Error tipo I”}) &= P(\text{“Rechazar } H_0\text{”} \mid \text{“}H_0 \text{ es verdadera”}) \\ &= P(\omega \in \mathscr{C} \mid \text{“}H_0 \text{ es verdadera”}) \\ &= P(\omega \in \{0,2,3\} \mid \text{“Se usó el dado”}) \\ &= 0 + 1/6 + 1/6 = 1/3. \end{aligned}$$

Por otro lado,

$$\begin{aligned} P(\text{“Error tipo II”}) &= P(\text{“No rechazar } H_0\text{”} \mid \text{“}H_0 \text{ es falsa”}) \\ &= P(\omega \notin \mathscr{C} \mid \text{“}H_0 \text{ es falsa”}) \\ &= P(\omega \in \{1,4,5,6\} \mid \text{“Se usaron las monedas”}) \\ &= 5/32 + 5/32 + 1/32 + 0 = 11/32. \end{aligned}$$

Naturalmente, si se cambia la regla de decisión cambiando la región de rechazo, cambian las probabilidades de los errores. En un problema de decisión de este tipo se busca encontrar una regla de decisión razonable que tenga errores menores. Vamos ahora a estudiar pruebas de hipótesis en el contexto de la estimación de parámetros en las distribuciones de probabilidad.

Definición. Una *hipótesis estadística* o simplemente *hipótesis* es una afirmación o conjetura acerca de la distribución de una o mas variables aleatorias.

Por ejemplo, si X tiene una distribución $\text{bin}(n,p)$, entonces la afirmación “p =0.2” es una hipótesis. Si X tiene una distribución $\text{N}(\mu,\sigma^2)$, entonces la afirmación “$\mu > 0$” es otro ejemplo de hipótesis estadística.

Una hipótesis es *simple* si especifica por completo la distribución de probabilidad en cuestión. Por ejemplo, si X tiene una distribución $\exp(\lambda)$, entonces la afirmación “$\lambda = 5$” es una hipótesis simple. Si X tiene una distribución $\text{N}(\mu,1)$, entonces la afirmación “$\mu = 0$” es otro ejemplo de hipótesis simple. En contraste, decimos que una hipótesis estadística es *compuesta* cuando no especifica por completo la distribución de probabilidad en cuestión. Si X tiene una distribución Poisson(λ), entonces “$\lambda > 20$” es una hipótesis compuesta. Si X tiene una distribución $\chi^2(n)$, entonces “$n \neq 5$” es otro ejemplo de una hipótesis compuesta. En general, contrastaremos dos hipótesis de acuerdo al siguiente esquema y notación:

$$H_0 : \text{(hipótesis nula)} \quad vs \quad H_1 : \text{(hipótesis alternativa)}.$$

Tanto la hipótesis nula (H_0) como la hipótesis alternativa (H_1) pueden ser simple o compuesta. De este modo tenemos cuatro diferentes tipos de contraste de hipótesis: simple vs simple, simple vs compuesta, compuesta vs simple, y compuesta vs compuesta.

Definición. Una *prueba de hipótesis* es una regla para decidir si se acepta la hipótesis nula o se rechaza en favor de la hipótesis alternativa.

Como sabemos, al tomar una decisión de este tipo se corre el riesgo de cometer errores. Al rechazo de la hipótesis nula cuando ésta es verdadera se le conoce como *error tipo I*, y a la probabilidad de cometer este primer tipo de error se le denota por la letra α. En cambio, a la aceptación de la hipótesis nula cuando ésta es falsa recibe el nombre de *error tipo II*, y a la probabilidad de cometer este segundo tipo de error se le denota por la letra β. Estas definiciones de errores se resumen en la siguiente tabla.

	H_0 cierta	H_0 falsa
Rechazar H_0	Eror tipo I con probabilidad α	Decisión correcta
No rechazar H_0	Decisión correcta	Error tipo II con probabilidad β

La información para obtener una regla de decisión que nos lleve a rechazar o no rechazar un hipótesis estadística provendrá de una muestra aleatoria $X_1, \ldots, X_n$ de la distribución de que se trate. Observe además que al aceptar una hipótesis no se afirma que ésta sea absolutamente cierta, sino simplemente que es consistente con los datos de la muestra aleatoria. Si la muestra cambia, posiblemente la decisión de rechazar o no rechazar también.

Definición. Se le llama *región crítica* a la región de rechazo de H_0, y a la probabilidad de cometer el error tipo I, esto es α, se le llama *tamaño de la región crítica*. A esta probabilidad se le conoce también con el nombre de *nivel de significancia*.

Nos limitaremos a mostrar la forma en la que se deriva la regla para llevar a cabo una prueba de hipótesis acerca de la media de una distribución normal.

Prueba acerca de la media de una distribución normal. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal con media desconocida μ y varianza conocida σ^2. Sabemos que $\bar{X}$ tiene distribución $\mathrm{N}(\mu, \sigma^2/n)$. Por lo tanto

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Sea μ_0 un número real particular. Deseamos contrastar las hipótesis $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$. El problema es encontrar una regla para decidir cuándo rechazar H_0 en favor de H_1 con base en los datos de la muestra $X_1, \ldots, X_n$. Cuando H_0 es cierta, esto es, cuando μ es efectivamente μ_0, tenemos que $\bar{X} \sim N(\mu_0, \sigma^2/n)$ y por lo tanto

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1).$$

La estadística $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ es una medida natural de la distancia entre $\bar{X}$, un estimador de μ, y su valor esperado μ_0 cuando H_0 es cierta. Es entonces razonable rechazar H_0 cuando la variable Z sea grande. Es por ello que tomamos como criterio de decisión rechazar H_0 cuando $|Z| \geq k$, para cierta constante k. ¿Cómo encontramos el número k? En una tabla de la distribución normal podemos encontrar un valor $z_{\alpha/2}$ tal que $P(|Z| \geq z_{\alpha/2}) = \alpha$, en donde α lo determina la persona que lleva a cabo la prueba de hipótesis, típicamente $\alpha = 0.1$. Véase la Figura 2.5. Este valor $z_{\alpha/2}$ es precisamente la constante k buscada pues con ello se logra que la región de rechazo sea de tamaño α.

A la variable aleatoria Z se le llama la *estadística de la prueba*, y la prueba se denomina *prueba de dos colas* pues la región de rechazo consta de las dos colas de la distribución normal que se muestran en la Figura 2.5. Llevar a cabo esta prueba de hipótesis consiste en usar los datos de la muestra para encontrar el valor de Z, si $|Z| \geq z_{\alpha/2}$, entonces se rechaza H_0, en caso contrario no se rechaza H_0. En la siguiente tabla se muestra resumida la información de esta prueba.

Prueba:	$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$
Estadística de prueba:	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$
Región de rechazo:	$\|Z\| \geq z_{\alpha/2}$
Error tipo I:	α
Error tipo II:	$\Phi(z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}})$, $\mu_1 \neq \mu_0$.

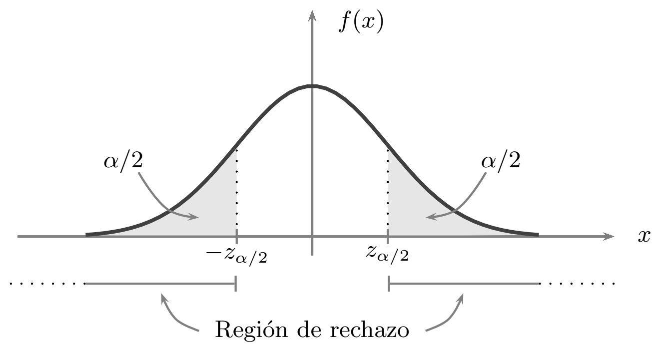


Figura 2.5:

Vamos a comprobar la fórmula que aparece en la tabla anterior acerca del error tipo II. Sea μ_1 cualquier número tal que $\mu_1 \neq \mu_0$. Calcularemos la probabilidad del error tipo II dado que el verdadero valor de μ es μ_1.

$$\begin{aligned}
\beta(\mu_1) &= P(\text{ “No rechazar } H_0 \text{ cuando } \mu = \mu_1\text{” }) \\
&= P(\,|Z| < z_{\alpha/2} \mid \mu = \mu_1\,) \\
&= P(\,|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\,| \; < z_{\alpha/2} \,|\, \mu = \mu_1\,) \\
&= P(\,\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \mid \mu = \mu_1\,) \\
&= P(\,-z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} < z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\,) \\
&= \Phi(z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}).
\end{aligned}$$

Ejercicio. En ciertas zonas de la ciudad se calcula el pago por el consumo de agua suponiendo un consumo promedio de 20,000 litros mensuales. Para verificar que si tal cantidad ha cambiado se han medido los consumos mensuales de 15 casas obteniéndose los siguientes resultados: 23456, 18325, 21982, 22371, 13292, 25073, 22601, 20930, 18788, 19162, 21442, 23935, 20320, 19095, 17421. ¿Debe cambiar el consumo promedio mensual para calcular los pagos o permanecer igual? Suponga $\sigma = 2000$. ∎

Puede también considerarse la prueba $H_0 : \mu = \mu_0$ contra $H_1 : \mu < \mu_0$ llamada *prueba de cola inferior* pues la región de rechazo consta de la cola izquierda de la distribución normal como se muestra en la Figura 2.6. Se rechaza la hipótesis H_0 sólo cuando los datos de la muestra son tales que $\bar{X}$ se encuentra muy a la izquierda de μ_0.

Prueba:	$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu < \mu_0$
Estadística de prueba:	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$
Región de rechazo:	$Z \leq -z_\alpha$
Error tipo I:	α
Error tipo II:	$1 - \Phi\left(-z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right), \ \mu_1 < \mu_0$

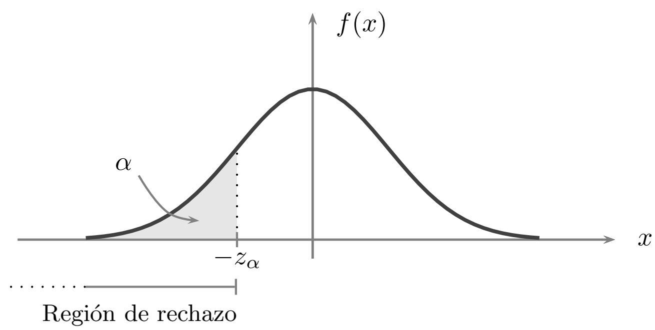


Figura 2.6:

Las características de la prueba $H_0 : \mu = \mu_0$ contra $H_1 : \mu > \mu_0$ llamada *prueba de cola superior* se muestran en la siguiente tabla, y la región de rechazo se presenta en la Figura 2.7. Se rechaza la hipótesis H_0 sólo cuando la muestra provee evidencia de que $\bar{X}$ se encuentra muy a la derecha de μ_0.

Prueba:	$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu > \mu_0$
Estadística de prueba:	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$
Región de rechazo:	$Z \geq z_\alpha$, (prueba de cola superior)
en donde $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$.Error tipo I:	α
Error tipo II:	$\Phi\left(z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right), \ \mu_1 > \mu_0.$

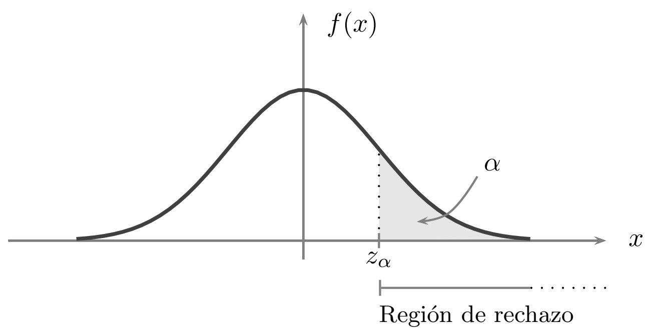


Figura 2.7:

Apéndice A

Ejercicios

Espacio muestral y eventos

1. Determine un espacio muestral para cada uno de los siguientes experimentos aleatorios.

   a) Lanzar una moneda tres veces.

   b) Lanzar a un mismo tiempo tres monedas indistinguibles.

   c) Escoger un número real al azar dentro del intervalo $[-1, 1]$ y después elevarlo al cuadrado.

   d) Registrar el número de llamadas telefónicas que llegan a un conmutador en un minuto dado.

   e) Colocar al azar dos bolas distinguibles en cuatro celdas numeradas.

   f) Colocar al azar dos bolas indistinguibles en cuatro celdas numeradas.

   g) Observar el marcador final de un juego de fútbol soccer.

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

3. Determine un experimento aleatorio para los siguientes espacios muestrales.

   a) $\Omega = \{0, 1, 2, \ldots\}$.

   b) $\Omega = [0, \infty)$.

   c) $\Omega = \{0, 1\}$.

Operaciones con conjuntos

4. Demuestre las leyes distributivas:
   a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
   b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
5. Demuestre las leyes de De Morgan:
   a) $(A \cup B)^c = A^c \cap B^c$.
   b) $(A \cap B)^c = A^c \cup B^c$.
6. Enuncie y demuestre las leyes de De Morgan para tres conjuntos. Para la demostración use la validez del mismo resultado para dos conjuntos.
7. Demuestre que $A = (A \cap B) \cup (A \cap B^c)$.
8. Demuestre que $A \cap B = B \cap (A \cup B^c)$.
9. Demuestre que $A \cap B = A \cap (B \cup A^c)$.
10. Demuestre que $A \cup B = B \cup (A \cap B^c)$.
11. Demuestre que $A \cup B = A \cup (B \cap A^c)$.
12. Demuestre que
    a) $A - B = A - (A \cap B)$.
    b) $A - B = (A \cup B) - B$.
13. Demuestre que
    a) $A \cap (B - C) = (A \cap B) - (A \cap C)$.
    b) $(A - C) \cap (B - C) = (A \cap B) - C$.
14. Demuestre que las siguientes dos definiciones de la operación diferencia simétrica son equivalentes.
    a) $A \triangle B = (A - B) \cup (B - A)$.
    b) $A \triangle B = (A \cup B) - (A \cap B)$.
15. Compruebe gráficamente las siguientes propiedades básicas de la diferencia simétrica.

a) $A\triangle(B\triangle C) = (A\triangle B)\triangle C$.

b) $A\triangle\emptyset = A$.

c) $A\triangle\Omega = A^c$.

d) $A\triangle A = \emptyset$.

e) $A\triangle A^c = \Omega$.

16. Sean $A = \{x \in \mathbb{R} : |x-2| \leq 4\}$ y $B = \{x \in \mathbb{R} : |x-1| > 2\}$. Muestre gráficamente los conjuntos A, B, A^c, B^c, $A\cap B$, $A\cup B$, $A-B$, $B-A$ y $A\triangle B$.

17. Sean $A = \{x \in \mathbb{R} : |x+1| \leq 2\}$ y $B = \{x \in \mathbb{R} : x^2 > 2\}$. Muestre gráficamente los conjuntos A, B, A^c, B^c, $A\cap B$, $A\cup B$, $A-B$, $B-A$ y $A\triangle B$.

18. Sean $A = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ y $B = \{(x,y) \in \mathbb{R}^2 : y > x^2\}$. Muestre gráficamente los conjuntos A^c, B^c, $A\cap B$, $A\cup B$, $A-B$, $B-A$ y $A\triangle B$.

19. Sean $A = \{(x,y) \in \mathbb{R}^2 : 2x^2 + 3y^2 \leq 6\}$ y $B = \{(x,y) \in \mathbb{R}^2 : y > -x^2\}$. Muestre gráficamente los conjuntos A^c, B^c, $A\cap B$, $A\cup B$, $A-B$, $B-A$ y $A\triangle B$.

20. *Ellos.* Si el 70 % de los hombres son feos, el 60 % son fuertes, y el 50 % son formales. ¿Cuál es el mínimo y el máximo posibles de hombres que son feos, fuertes y formales?

21. *Ellas.* Si el 90 % de las mujeres son bonitas, el 80 % son sensibles, y el 70 % son sinceras. ¿Cuál es el mínimo y el máximo posibles de mujeres que son bonitas, sensibles y sinceras?

Conjuntos ajenos

22. Compruebe que los conjuntos A y $B-A$ son siempre ajenos.

23. ¿Son los conjuntos $\{(x,y) : y = e^{-x}\}$ y $\{(x,y) : y = 1-x\}$ ajenos? En caso negativo, ¿Cuántos puntos en común tienen?

24. Sean A y B ajenos, y sea $A_1 \subseteq A$. Compruebe que A_1 y B son ajenos. Intuitivamente la afirmación es evidente, el reto consiste en encontrar una justificación matemática de ello.

25. ¿Cuántas igualdades son necesarias verificar en general para comprobar que n conjuntos son ajenos dos a dos?

Conjunto potencia

26. ¿Quién es $2^{\emptyset}$?

27. Demuestre que si $A \subseteq B$, entonces $2^A \subseteq 2^B$.

Producto Cartesiano

28. Sean $A = \{a_1, a_2\}$, $B = \{b_1, b_2, b_3\}$ y $C = \{c_1, c_2\}$. Escriba explícitamente a los elementos del producto Cartesiano $A \times B \times C$.

29. Sea $\Omega = \{a, b, c\}$. Escriba explícitamente al conjunto $\Omega \times \Omega$.

Probabilidad

30. Compruebe que la definición de probabilidad clásica cumple con los tres axiomas de la probabilidad.

31. Compruebe que la definición de probabilidad frecuentista cumple con los tres axiomas de la probabilidad.

32. Sean P_1 y P_2 dos medidas de probabilidad definidas sobre la misma clase de subconjuntos de Ω. Sea α un número real en el intervalo $[0, 1]$. Demuestre que $P = \alpha P_1 + (1 - \alpha) P_2$ satisface los tres axiomas de la probabilidad.

33. Demuestre que $P(\emptyset) = 0$, sin usar $P(\Omega) = 1$.

34. Demuestre que $P(B - A) = P(B) - P(A \cap B)$, para cualesquiera eventos A y B.

35. Demuestre que $0 \leq P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B) \leq 2$.

36. Demuestre que $P(A \cap B) \geq P(A) + P(B) - 1$.

37. Demuestre que $P(A \triangle B) = P(A) + P(B) - 2P(A \cap B)$. Esta es la probabilidad de que suceda exactamente uno de los eventos A y B.

38. Demuestre que $P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A \cap B)$.

39. Sean A y B eventos ajenos tales que $P(A)$ =0.3 y $P(B)$ =0.2. Encuentre

a) $P(A \cup B)$.

b) $P(A^c)$.

c) $P(A^c \cap B)$.

d) $P(A \cap B^c)$.

e) $P(A \triangle B)$.

40. Diga falso o verdadero. Justifique en cada caso.

a) Si $P(A) = 0$, entonces $P(A \cap B) = 0$.

b) Si $P(A) = P(B)$, entonces $A = B$.

c) Si $P(A) \leq P(B)$, entonces $A \subseteq B$.

41. Diga falso o verdadero. Justifique en cada caso.

a) Si $P(A) > 0$, entonces $P(A \cup B) > 0$.

b) Si $P(A) > 0$, entonces $P(A \cap B) > 0$.

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

d) Si $P(A) > 0$, entonces $P(A^c) > 0$.

42. Diga falso o verdadero. Justifique en cada caso.

a) $P(B - A) = P(B) - P(A)$.

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

c) Si $P(A) > 1/2$, entonces $P(A^c) < 1/2$.

43. La probabilidad de un cierto evento es tal que duplica la probabilidad de su complemento. ¿Cuánto vale la probabilidad de este evento?

Análisis combinatorio

44. ¿De cuántas formas distintas pueden seis personas formarse en una fila lineal?

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

46. ¿Cuántas diagonales se pueden trazar en un polígono convexo de n lados?

47. ¿De cuántas maneras diferentes pueden clasificarse los tres primeros lugares de una carrera de 15 corredores ?

48. ¿Cuántos enteros positivos de a lo mas cinco dígitos son divisibles por 2? ¿Cuántos hay que empiecen con el dígito 1?

49. ¿Cuántos equipos de 3 personas se pueden formar de un grupo de 5 personas?

50. Demuestre que $\binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1}$.

51. Demuestre que $\binom{n}{k} = \frac{n}{n-k}\binom{n-1}{k}$.

52. Demuestre que $\binom{n-1}{k} = \frac{k+1}{n}\binom{n}{k+1}$.

53. Demuestre que $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$. Esta es la fórmula para construir el triángulo de Pascal.

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

55. ¿Cuántos números binarios diferentes se pueden obtener al usar los siete dígitos 1010101 ?

56. ¿Cuántas "palabras" diferentes se pueden formar con todos los caracteres (incluyendo repeticiones) de la palabra AMAR?

57. *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.

58. Encuentre el total de números enteros de cuatro dígitos sin repetición, tomados del conjunto $\{0, 1, 2, \ldots, 9\}$ de tal manera que ningún número empiece con 0. ¿Cuántos de ellos son pares y cuántos son impares?

59. En una clase de 25 estudiantes hay 15 mujeres y 10 hombres. a) ¿De cuántas formas puede seleccionarse un comité de tres hombres y tres mujeres ? b) ¿De cuántas formas puede seleccionarse un comité de seis estudiantes? c) ¿De cuántas formas puede seleccionarse un comité de seis estudiantes todos del mismo sexo?

60. ¿De cuántas formas diferentes se pueden colocar los números $1, 2, 3, 4, 5$ y 6 en las seis caras de un dado? Observe que originalmente las caras del dado son indistinguibles.

61. Demuestre que $\#(2^{\Omega}) = 2^{\#\Omega}$. Sugerencia: Use el método de inducción sobre $n = \#\Omega$.

62. Tenemos n jugadores finalistas en un torneo de ajedrez. Para escoger al ganador se establece que cada jugador debe jugar contra cada uno de los otros finalistas.

    a) ¿Cuántas partidas se llevarán a cabo?

    b) Suponiendo que no hay empates y que cada jugador gana un punto por cada juego ganado, ¿De cuántas formas distintas se pueden asignar la totalidad de puntos en el conjunto de jugadores?

    c) ¿Cuántas configuraciones finales existen en donde haya un único jugador con mayor número de puntos?

63. Demuestre que el número máximo de regiones en las que n líneas rectas dividen un plano es $1 + n(n + 1)/2$.

64. Sea A_n en número máximo de regiones en los que n planos dividen el espacio. Demuestre que $A_{n+1} = A_n + 1 + n(n + 1)/2$.

65. Demuestre que el número máximo de regiones en los que n círculos dividen el plano es 2^n.

66. Sea A_n en número máximo de regiones en los que n esferas dividen el espacio. Demuestre que $A_{n+1} = A_n + n^2 - n + 2$.

67. ¿Cuántos divisores diferentes tiene el número $5^3 \cdot 7^4$?

68. Se asignan 40 problemas para un examen de probabilidad. El examen consistirá de 4 problemas escogidos al azar, y cada problema tendrá un peso de 25 puntos. Si un alumno resuelve únicamente 20 de los problemas asignados, ¿Cuál es la probabilidad de que el alumno obtenga

    a) cero puntos?

    b) 25 puntos?

    c) 50 puntos?

    d) 75 puntos?

e) 100 puntos?

69. ¿Cuántos subconjuntos podemos obtener de un conjunto de n elementos? Escriba explícitamente todos los subconjuntos del conjunto $\{a, b, c, d\}$.

70. Sea E el conjunto de vectores de $\mathbb{R}^n$ para los cuales alguna de sus coordenadas es cero. ¿Cuántas regiones conexas tiene el conjunto $\mathbb{R}^n - E$?

71. ¿Cuántas configuraciones diferentes se pueden obtener en un tablero de ajedrez después de

    a) la primera jugada del primer jugador?

    b) la primera jugada del segundo jugador?

72. Use el método de inducción para demostrar el teorema del binomio,

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

73. En una sala de cómputo hay seis computadoras numeradas del 1 al 6. ¿De cuántas formas distintas pueden las seis computadoras estar siendo usadas o no usadas?

74. ¿De cuántas formas posibles se puede ordenar el conjunto $\{1, 2, \ldots, 2n+1\}$ de tal forma que cada número impar ocupe una posición impar?

75. ¿De cuántas formas posibles se puede ordenar el conjunto $\{1, 2, \ldots, 2n\}$ de tal forma que cada número par ocupe una posición par?

76. ¿Cuántas "palabras" diferentes podemos obtener usando todas las letras (incluyendo repeticiones) de la palabra "manzana"?

77. ¿Cuántas "palabras" diferentes podemos obtener usando todas las sílabas (incluyendo repeticiones) de la palabra "cucurrucucú"?

78. Sea $n \geq 1$ un entero. ¿Cuántas soluciones enteras no negativas tiene la ecuación

    a) $x_1 + x_2 + \cdots + x_k = n$?

    b) $x_1 + x_2 + \cdots + x_k \leq n$?

    c) $x_1 + x_2 + \cdots + x_k \geq n$?

79. Sean $n \geq k$ dos enteros positivos. ¿Cuántos vectores con entradas enteras no negativas $(x_1, x_2, \ldots, x_k)$ satisfacen las restricciones

$$1 \leq x_1 < x_2 < \cdots < x_k \leq n\,?$$

80. Desarrolle la expresión $(a+b+c)^3$ usando la fórmula de coeficientes multinomiales (1.1) en la página 26, y después compruebe la fórmula directamente multiplicando el trinomio por si mismo.

Probabilidad condicional e independencia

81. Sean A y B dos eventos independientes. Demuestre que

   a) A^c y B son independientes.

   b) A y B^c son independientes.

   c) A^c y B^c son independientes.

82. Sean A y B eventos independientes tales que $P(A) = 0.1$ y $P(B) = 0.5$. Encuentre

   a) $P(A \cap B)$.
   b) $P(A^c \cap B)$.
   c) $P(A \cap B^c)$.
   d) $P(A^c \cap B^c)$.
   e) $P(A \cup B)$.
   f) $P(A^c \cup B)$.
   g) $P(A \cup B^c)$.
   h) $P(A^c \cup B^c)$.

83. Sea P una medida de probabilidad y B un evento con probabilidad positiva. Demuestre que la probabilidad condicional $P(\cdot|B)$ satisface los tres axiomas de la probabilidad.

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

$$P(A \cap B \cap C) = p^3.$$

85. Diga falso o verdadero. Justifique en cada caso.

   a) El número $P(A|B)$ nunca es cero.

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

   c) $P(A|B) \leq P(A)$.

86. Diga falso o verdadero. Justifique en cada caso.

    a) $P(A|B) + P(A^c|B) = 1$.
    b) $P(A|B) + P(A|B^c) = P(A)$.
    c) $P(A|A \cap B) = P(B|A \cap B) = 1$.

87. Diga falso o verdadero. Justifique en cada caso.

    a) $P(\Omega|B) = 1$.
    b) $P(\emptyset|B) = 0$.
    c) $P(A|A) = P(A)$.

88. Diga falso o verdadero. Justifique en cada caso.

    a) $P(A|B_1 \cup B_2) = P(A|B_1) + P(A|B_2)$ cuando B_1 y B_2 son ajenos.
    b) $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$ cuando A_1 y A_2 son ajenos.
    c) $P(A|B_1 \cap B_2) = P(A|B_1)P(A|B_2)$.
    d) $P(A_1 \cap A_2|B) = P(A_1|B)P(A_2|B)$.

89. Demuestre que

    a) el conjunto $\emptyset$ es independiente consigo mismo.
    b) el conjunto Ω es independiente consigo mismo.

90. Diga falso o verdadero. Justifique en cada caso.

    a) A, B independientes $\Rightarrow A, B$ ajenos.
    b) A, B ajenos $\Rightarrow A, B$ independientes.

91. Sea A cualquier evento. Demuestre que A y $\emptyset$ son eventos independientes.

92. Sea A cualquier evento. Demuestre que A y Ω son eventos independientes.

93. Sean A y B eventos independientes. Demuestre que

$$P(A \cup B) = 1 - P(A^c)P(B^c).$$

94. Diga falso o verdadero. Justifique en cada caso.

    a) A, B independientes $\Rightarrow B, A$ independientes.
    b) Para cualquier evento A, A es independiente con A.

c) A, B independientes y B, C independientes $\Rightarrow A, C$ independientes.

95. Sean A y B eventos independientes tales que $P(A) = p_1$ y $P(B) = p_2$. Calcule la probabilidad de que no ocurra ninguno de estos eventos.

96. Sean A y B eventos independientes y ajenos. Demuestre que forzosamente alguno de estos eventos tiene probabilidad cero.

97. Considere un experimento aleatorio con espacio muestral $\Omega = (0, 1)$. Definimos la probabilidad de un intervalo $(a, b) \subseteq (0, 1)$ 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.

98. ¿Cuántas igualdades son necesarias verificar para demostrar que n eventos son independientes? Justifique su respuesta.

Teorema de probabilidad total

99. *La urna de Polya*. Suponga que en una urna se tienen b bolas blancas y r bolas rojas. Un experimento aleatorio 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". Compruebe que para cada $n \geq 1$,
$$P(R_n) = \frac{r}{r+b}.$$

100. Una persona toma al azar, con idéntica probabilidad, uno de los números $1, 2$ o 3, y luego tira un dado equilibrado tantas veces como indica el número escogido. Después suma el resultado de las tiradas del dado. ¿Cuál es la probabilidad de que obtenga un total de 5?

Teorema de Bayes

101. En una urna se encuentran b bolas de color blanco y n bolas de color negro. A un mismo tiempo se extraen k bolas al azar y resulta que todas son del mismo color. ¿Cuál es la probabilidad de que las bolas sean de color negro?

102. Se escoge al azar un dígito binario para ser enviado a través de un canal de transmisión. Se escoge el "0" con probabilidad 0.4, y se escoge el "1" con probabilidad 0.6. El canal de comunicación es ruidoso de modo que un "0" se distorsiona en un "1" con probabilidad 0.2, y un "1" se distorsiona en un "0" con probabilidad 0.1. Encuentre la probabilidad de que

    a) se reciba un "0".
    b) se reciba un "1".
    c) se haya enviado un "0" dado que se recibió un "0".
    d) se haya enviado un "1" dado que se recibió un "1".

Variables aleatorias

103. Determine en cada caso si la variable aleatoria en cuestión es discreta o continua. ¿Cuáles son sus posible valores?

    a) Tiempo de vida de una persona escogida al azar.
    b) Número de errores tipográficos en una página escogida al azar de un libro.
    c) Tiempo de servicio en una transacción escogida al azar realizada por una persona en un cajero automático.
    d) Monto de una reclamación por accidente automovilístico escogida al azar del conjunto de reclamaciones efectuadas a una compañía aseguradora.

104. Considere el experimento aleatorio de escoger un número al azar dentro del intervalo unitario $(0, 1)$. Suponga que cada resultado de este experimento se escribe en su expansión decimal como $\omega = 0.x_1x_2x_3\cdots$ Determine en los siguientes casos el conjunto de valores de la variable aleatoria definida y clasifique ésta como discreta o continua.

    a) $X(\omega) = \omega$.
    b) $X(\omega) = x_1$.
    c) $X(\omega) = 1 - \omega$.
    d) $X(w) = 0.0x_1x_2x_3\cdots$

105. Considere un experimento aleatorio con espacio muestral equiprobable $\Omega = \{1,2,3,4,5,6\}$. Defina la variable aleatoria $X(\omega) = 2(\omega - 3)$. ¿Cuáles son los posibles valores de X? Calcule $P(X = 0)$, $P(X \in \{2,3\})$, $P(X \geq 0)$, $P(X < 0)$, $P(X^2 = 1)$, $P(2X - 4 = 0)$, y $P(X^2 = 4)$.

106. Considere el ejemplo del experimento aleatorio de lanzar un dardo en un tablero circular de radio uno, Figura 1.13, junto con las variables aleatorias $X(x,y) = x$, $Y(x,y) = y$ y $Z(x,y) = \sqrt{x^2+y^2}$. Suponga que para cada región $A \subseteq \Omega$ cuya área puede ser calculada se define $P(A) = \text{Área}(A)/\text{Área}(\Omega)$. Calcule $P(X \geq 0)$, $P(X < 0)$, $P(Y > X)$, $P(X + Y \leq 1)$, $P(Z < 1/2)$ y $P(1/3 < Z < 1/2)$.

Funciones de densidad y de distribución

107. Encuentre el valor de la constante c para que $f(x)$ sea una función de probabilidad. Grafique esta función y calcule $P(X \in \{2,3,4\})$ y $P(X < 3)$ en cada caso.

 a) $f(x) = \begin{cases} c\,x & \text{si } x = 1,2,\ldots,10, \\ 0 & \text{otro caso.} \end{cases}$

 b) $f(x) = \begin{cases} c\,x^2 & \text{si } x = 1,2,\ldots,10, \\ 0 & \text{otro caso.} \end{cases}$

108. Encuentre el valor de la constante c para que la siguiente función sea de densidad. Grafique $f(x)$ y calcule $P(X \geq \pi)$ y $P(X \in [-\pi, 2\pi])$.

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

109. Encuentre el valor de la constante c para que la siguiente función sea de densidad. Grafique $f(x)$ y calcule $P(X \in (1,\infty))$.

$$f(x) = c\,e^{-|x|}, \quad \text{para } -\infty < x < \infty.$$

110. Explique porqué no es posible encontrar un valor de la constante c para que la siguiente función sea de probabilidad o de densidad.

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

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

111. Sea X discreta con función de probabilidad dada por la siguiente tabla. Grafique $f(x)$ y calcule $P(X \geq 0)$, $P(X < 0)$ y $P(X^2 = 1)$.

x	-1	0	1
$f(x)$	0.2	0.3	0.5

112. Sea X discreta con función de probabilidad dada por la tabla que aparece abajo. Grafique $f(x)$. Calcule la función de probabilidad de las siguientes variables aleatorias $Y = X^2$, $Z = |X|$ y $W = 2X - 5$. Grafique en cada caso.

x	-2	-1	0	2	3	5
$f(x)$	0.1	0.15	0.4	0.1	0.15	0.1

113. Sea X discreta con función de densidad dada por la tabla que aparece abajo. Encuentre el valor de c y grafique $f(x)$. Calcule y grafique la función de probabilidad de la variable $Y = X^2$.

x	-2	0	2
$f(x)$	0.1	c	0.1

114. Sea X continua con función de densidad

$$f(x) = \begin{cases} 1/10 & \text{si } -k \leq x \leq 4k \\ 0 & \text{otro caso.} \end{cases}$$

a) Determine el valor de la constante k y grafique $f(x)$.
b) Calcule y grafique $F(x)$.
c) Calcule $P(-1 \leq X \leq 3)$, $P(X \geq 2)$ y $P(X \leq 0)$.
d) Encuentre m tal que $P(|X - 1| \geq m) = 1/2$.

115. Sea X una variable aleatoria con la función de distribución que aparece abajo. ¿Es X discreta o continua? Grafique $F(x)$. Encuentre y grafique la correspondiente función de densidad $f(x)$. Calcule además $P(X = 2)$ y $P(1 < X < 2)$.

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

116. Sea X una variable aleatoria con la función de distribución que aparece abajo. ¿Es X discreta o continua? Grafique $F(x)$. Encuentre y grafique la correspondiente función de densidad $f(x)$. Calcule además $P(X = 1/2)$ y $P(X > 1/2)$.

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

117. Una urna contiene cuatro bolas numeradas $1, 2, 3$ y 4. Se extraen dos bolas al azar, una a la vez y sin reemplazo. Sea X la variable aleatoria que denota la suma de los números de las dos bolas seleccionadas.

    a) Determine Ω.
    b) Calcule y grafique $f(x)$.
    c) Calcule y grafique $F(x)$.
    d) Calcule $P(X \geq 6)$, $P(3 < X \leq 5)$ y $P(X = 6)$.

118. Determine si la siguiente función es de probabilidad. Grafique la función y justifique su respuesta.

$$f(x) = \begin{cases} 1/6 & \text{si } x = 0, 1, \\ 2/3 & \text{si } x = 2, \\ 0 & \text{otro caso.} \end{cases}$$

119. Determine si la siguiente función es de probabilidad. Grafique la función y justifique su respuesta.

$$f(x) = \begin{cases} \binom{4}{x} \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{4-x} & \text{si } x = 0, 1, 2, 3, 4, \\ 0 & \text{otro caso.} \end{cases}$$

120. Determine si cada una de las siguientes funciones es de densidad. Grafique la función en cada caso y justifique su respuesta.

    a) $f(x) = \begin{cases} 4x/5 & \text{si } x \in [0, 2], \\ 0 & \text{otro caso.} \end{cases}$

    b) $f(x) = \begin{cases} 2x^2/3 - 2x + 4/3 & \text{si } x \in [0, 3], \\ 0 & \text{otro caso.} \end{cases}$

121. Sea X una variable aleatoria con la siguiente función de distribución. Encuentre y grafique $f(x)$. Calcule $P(0 \leq X < 10)$.

$$F(x) = \begin{cases} 1 - (1/2)^{x+1} & \text{si } x = 0, 1, 2, 3, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

122. Sea X una variable aleatoria continua con la función de densidad que aparece abajo. Encuentre el valor de la constante c y grafique la función $f(x)$. Encuentre y grafique además la función de distribución $F(x)$.

$$f(x) = \begin{cases} 2x/9 & \text{si } 0 < x < c, \\ 0 & \text{otro caso.} \end{cases}$$

Esperanza, varianza, momentos

123. Sea a un número fijo. Construya una variable aleatoria X tal que $E(X) = a$.

124. Calcule la esperanza de la variable aleatoria discreta X cuya función de probabilidad es

  a) $f(x) = \begin{cases} 1/3 & \text{si } x = 0, 1, \\ 1/6 & \text{si } x = 2, 3, \\ 0 & \text{otro caso.} \end{cases}$

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

125. Calcule la esperanza de la variable aleatoria continua X cuya función de densidad es

  a) $f(x) = e^{-x}$, para $x > 0$.

  b) $f(x) = 6x(1 - x)$, para $0 < x < 1$.

126. Sea X una variable aleatoria discreta con la función de probabilidad que aparece abajo. Demuestre que $f(x)$ es efectivamente una probabilidad de densidad y que la esperanza de X no existe. Este es un ejemplo de una variable aleatoria discreta que no tiene esperanza finita.

$$f(x) = \begin{cases} \dfrac{1}{x(x+1)} & \text{si } x = 1, 2, 3, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

127. Sea X una variable aleatoria continua con la función de densidad que aparece abajo. Demuestre que esta función es efectivamente una función de densidad. Compruebe además que la esperanza de X no existe. Este es un ejemplo de una variable aleatoria continua que no tiene esperanza finita. Es una caso particular de la distribución Cauchy.

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad \text{para} -\infty < x < \infty.$$

128. Diga falso o verdadero. Justifique en cada caso.

    a) La esperanza de una v.a. puede ser cero.
    b) No hay dos v.a.s distintas con la misma esperanza.
    c) La esperanza de una v.a. nunca es negativa.
    d) La varianza de una v.a. puede ser cero.
    e) La varianza de una v.a. nunca es negativa.
    f) No hay dos v.a.s distintas con la misma varianza.

129. Demuestre que

    a) $E(E(X)) = E(X)$.
    b) $\text{Var}(\text{Var}(X)) = 0$.

130. Sea X la variable aleatoria constante c. Compruebe que
    a) $E(X) = c$. b) $E(X^n) = c^n$. c) $\text{Var}(X) = 0$.

131. Calcule la media y varianza de la variable aleatoria X con función de probabilidad

$$f(x) = \begin{cases} 1/9 & \text{si } x = 0, 1, 2, \\ 2/9 & \text{si } x = 3, 4, 5, \\ 0 & \text{otro caso.} \end{cases}$$

132. Calcule la media y varianza de la variable aleatoria X cuya función de probabilidad es

$$f(x) = \begin{cases} (1/2)^{x+1} & \text{si } x = 0, 1, 2, 3, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

133. Diga falso o verdadero. Justifique en cada caso.

    a) $\text{Var}(E(X)) = 0$.

b) $E(\text{Var}(X)) = \text{Var}(X)$.

134. Sea X una variable aleatoria continua con función de densidad $f(x) = \frac{1}{2}e^{-|x|}$, para $-\infty < x < \infty$. Demuestre que $f(x)$ es efectivamente una función de densidad y compruebe que

a) $E(X) = 0$.

b) $E(X^2) = 2$.

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

d) $E(X^n) = n!$ para n par.

135. Diga falso o verdadero. Justifique en cada caso.

a) $E(-X) = -E(X)$.

b) $\text{Var}(-X) = -\text{Var}(X)$.

c) $E(\text{Var}(X)) = \text{Var}(E(X))$.

136. Encuentre el error en la siguiente "demostración" de la afirmación de que la varianza de cualquier variable aleatoria es cero.

$$\begin{aligned} 0 &= \text{Var}(0) \\ &= \text{Var}(X + (-X)) \\ &= \text{Var}(X) + \text{Var}(-X) \\ &= \text{Var}(X) + \text{Var}(X) \\ &= 2\,\text{Var}(X). \end{aligned}$$

Distribución uniforme discreta

137. Sea X con distribución uniforme en el conjunto $\{1, \ldots, n\}$. Demuestre que

a) $E(X) = (n + 1)/2$.

b) $E(X^2) = (n + 1)(2n + 1)/6$.

c) $\text{Var}(X) = (n^2 - 1)/12$.

138. Se escogen completamente al azar y de manera independiente dos números a y b dentro del conjunto $\{1, 2, \ldots, 9, 10\}$. ¿Cuál es la probabilidad de que el cociente a/b sea menor a uno?

Distribución Bernoulli

139. Sea X una variable aleatoria con distribución Ber(p). Verifique que $f(x)$ es efectivamente una función de densidad. Demuestre además que

    a) $E(X) = p$.

    b) $E(X^n) = p$, para $n \geq 1$.

    c) $\text{Var}(X) = p(1-p)$.

Distribución binomial

140. Sea X una variable aleatoria con distribución bin(n, p). Verifique que $f(x)$ es efectivamente una función de densidad.

141. Sea X una variable aleatoria con distribución bin(n, p). Demuestre que

    a) $E(X) = np$.

    b) $\text{Var}(X) = np(1-p)$.

142. Sea X una variable aleatoria con distribución bin(n, p) tal que $E(X) = 4$ y $\text{Var}(X) = 2$. ¿Cuáles son los valores de n y p?

143. Sea X una variable aleatoria con distribución bin(n, p). Demuestre que la variable $Y = n - X$ tiene distribución bin($n, 1-p$). Proporcione una explicación probabilísta de este resultado.

144. Sea X con distribución bin(n, p). Demuestre que para $x = 0, 1, \ldots, n-1$, se cumple la siguiente fórmula. Esta expresión permite calcular las probabilidades de esta distribución de una forma iterativa.

$$P(X = x + 1) = \frac{p}{(1-p)} \frac{(n-x)}{(x+1)} P(X = x).$$

145. Se lanza una moneda equilibrada 6 veces. Calcule la probabilidad de que cada cara caiga exactamente 3 veces.

146. Se lanza una moneda equilibrada $2n$ veces. Calcule la probabilidad de que ambas caras caigan el mismo número de veces.

147. Sea X una variable aleatoria con distribución bin(n, p). Demuestre que $0 \leq \text{Var}(X) \leq E(X)$.

148. Suponiendo que es igualmente probable que nazca un hombre (H) o una mujer (M), y considerando la observación de 6 nacimientos. ¿Cuál de los siguientes eventos es más probable que ocurra?

     a) MHHMHM   b) MMMMHM   c) HMHMHM

Distribución geométrica

149. Sea X una variable aleatoria con distribución geo(p). Verifique que la función $f(x)$ es efectivamente una función de densidad.

150. Sea X una variable aleatoria con distribución geo(p). Demuestre que

     a) $E(X) = (1-p)/p$.

     b) $\text{Var}(X) = (1-p)/p^2$.

151. Sea X una variable aleatoria con distribución geo(p). Demuestre que para cualesquiera $a, b = 0, 1, 2, \ldots$ se cumple la siguiente propiedad llamada de pérdida de memoria: $P(X \geq a + b \,|\, X \geq a) = P(X \geq b)$.

152. Considere una urna con 3 bolas negras y 5 bolas blancas. Se escoge una bola al azar, se registra su color, y después se regresa a la urna. ¿Cuántas extracciones en promedio se necesitan realizar hasta obtener una bola negra por primera vez?

Distribución Poisson

153. Sea X una variable aleatoria con distribución Poisson(λ). Verifique que $f(x)$ es efectivamente una función de densidad y demuestre que $E(X) = \lambda$, y $\text{Var}(X) = \lambda$.

154. Sea X una variable aleatoria con distribución Poisson(λ). Demuestre que para $x = 0, 1, 2, \ldots$ se cumple la siguiente fórmula. Esta expresión permite calcular las probabilidades Poisson de una forma iterativa.

$$P(X = x + 1) = \frac{\lambda}{(x+1)} P(X = x).$$

155. Sea X una variable aleatoria con distribución Poisson(λ). Demuestre que la probabilidad de que X tome un valor par es $(1 + e^{-2\lambda})/2$.

156. El número de computadoras que fallan por mes en un laboratorio de cómputo tiene una distribución Poisson con un promedio mensual de $\lambda = 2$ máquinas descompuestas. El laboratorio tiene capacidad para reparar hasta dos máquinas por mes. Cuando se descomponen mas de dos máquinas, las restantes se envían fuera del laboratorio para su reparación.

    a) ¿Cuál es la probabilidad de que en un mes cualquiera sea necesario enviar máquinas fuera del laboratorio para su reparación?

    b) Responda al inciso anterior cuando se reduce la capacidad de reparación del laboratorio a una computadora por mes.

    c) ¿Cuál es el número de computadoras con falla más probable en un mes?

Distribución binomial negativa

157. Sea X una variable aleatoria con distribución bin neg(r, p). Verifique que $f(x)$ es efectivamente una función de densidad.

158. Sea X una variable aleatoria con distribución bin neg(r, p). Demuestre que $E(X) = r(1-p)/p$, y $\mathrm{Var}(X) = r(1-p)/p^2$.

159. Sea X una variable aleatoria con distribución bin neg(r, p). Demuestre que para $x = 1, 2, \ldots$ se cumple la siguiente fórmula. Esta expresión permite calcular las probabilidades de esta distribución de una forma iterativa.

$$P(X = x) = (1-p)\,\frac{r + x - 1}{x}\,P(X = x - 1).$$

Distribución hipergeométrica

160. Sea X con distribución hipergeo(N, K, n). Verifique que $f(x)$ es efectivamente una función de probabilidad.

Distribución uniforme continua

161. Sea X con distribución unif(a, b). Verifique que $f(x)$ es efectivamente una función de densidad.

162. Sea X con distribución uniforme en el intervalo (a, b). Demuestre que

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

   b) $\text{Var}(X) = (b - a)^2/12$.

163. Sea X con distribución uniforme en el intervalo $(0, 1)$. Demuestre que el n-ésimo momento de X es $1/(n + 1)$.

164. Sea X con distribución uniforme en el intervalo (a, b). Demuestre que

$$E(X^n) = \frac{b^{n+1} - a^{n+1}}{(n + 1)(b - a)}.$$

165. Sea X una variable aleatoria con distribución uniforme en el intervalo $(-1, 1)$. Demuestre que

$$E(X^n) = \begin{cases} 1/(n + 1) & \text{si } n \text{ es par,} \\ 0 & \text{si } n \text{ es impar.} \end{cases}$$

166. Sea X con distribución unif$(0, 1)$. Obtenga la distribución de la variable $Y = 10X - 5$.

167. Sea X con distribución unif(a, b). Demuestre que la media y la mediana de X coinciden.

Distribución exponencial

168. Sea X una variable aleatoria con distribución $\exp(\lambda)$. Verifique que $f(x)$ es efectivamente una función de densidad.

169. Sea X con distribución $\exp(\lambda)$. Demuestre que

   a) $E(X) = 1/\lambda$.

   b) $\text{Var}(X) = 1/\lambda^2$.

170. Sea X una variable aleatoria con distribución $\exp(\lambda)$. Demuestre que para cualesquiera números $x, y \geq 0$ se cumple la igualdad que aparece abajo. La distribución exponencial es la única distribución continua que satisface esta propiedad llamada de *pérdida de memoria*.

$$P(X \geq x + y \,|\, X \geq x) = P(X \geq y).$$

171. Sea X con distribución $\exp(\lambda)$. Demuestre que para cualesquiera números $x, y \geq 0$, $F(x+y) = F(x) + F(y) - F(x)\,F(y)$.

172. Suponga que el tiempo que un usuario cualquiera permanece conectado a un servidor en una red de cómputo se puede modelar como una variable aleatoria con distribución exponencial con media igual a 10 minutos. De mil usuarios, ¿Cuántos tienen un conexión superior a una hora? Calcule además la probabilidad de que un usuario cualquiera

    a) no permanezca conectado mas de 10 minutos.

    b) permanezca conectado mas de 10 minutos pero menos de una hora.

Distribución gama

173. Sea X una variable aleatoria con distribución gama(n, λ). Verifique que $f(x)$ es efectivamente una función de densidad.

174. Sea X con distribución gama(n, λ). Demuestre que

    a) $E(X) = n/\lambda$.

    b) $\text{Var}(X) = n/\lambda^2$.

    c) $E(X^m) = \Gamma(m+n)/(\,\lambda^m \Gamma(n)\,)$.

175. Demuestre las siguientes propiedades de la función gama.

    a) $\Gamma(n+1) = n\,\Gamma(n)$.

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

    c) $\Gamma(n+1) = n!$ si n es entero.

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

Distribución beta

176. Sea X con distribución beta(a, b). Verifique que $f(x)$ es efectivamente una función de densidad.

177. Sea X con distribución beta(a, b). Demuestre que

    a) $E(X) = a/(a+b)$.

b) $\mathrm{Var}(X) = ab/(\,(a+b+1)(a+b)^2\,)$.

178. Demuestre las siguientes propiedades de la función beta.

a) $B(a,b) = B(b,a)$.

b) $B(a,1) = 1/a$.

c) $B(1,b) = 1/b$.

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

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

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

g) $B(1/2,1/2) = \pi$.

179. Sea X una variable aleatoria con distribución beta(a,b), con $a = b = 1$. Demuestre que X tiene una distribución uniforme en el intervalo $(0,1)$.

180. Sea X con distribución beta$(1/2,1/2)$. En este caso se dice que X tiene una *distribución arcoseno*.

a) Calcule y grafique $f(x)$.

b) Demuestre que $f(x)$ es una función de densidad.

c) Calcule $E(X)$ y $\mathrm{Var}(X)$.

181. Sea X con distribución beta(a,b). Demuestre que cuando $a > 0$ y $b = 1$, la función de distribución de X toma la forma

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

182. Sea X con distribución beta(a,b). Demuestre que cuando $a = 1$ y $b > 0$, la función de distribución de X toma la forma

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

183. Sea X con distribución beta(a,b). Demuestre que para cualquier entero $n \geq 0$, $E(X^n) = B(a+n,b)/B(a,b)$.

Distribución normal

184. Sea X con distribución $\mathrm{N}(\mu, \sigma^2)$. Verifique que $f(x)$ es efectivamente una función de densidad.

185. Sea X con distribución $\mathrm{N}(\mu, \sigma^2)$. Demuestre que $E(X) = \mu$, y $\mathrm{Var}(X) = \sigma^2$.

186. Sea X con distribución $\mathrm{N}(\mu, \sigma^2)$. Demuestre que la variable $Z = (X - \mu)/\sigma$ tiene distribución normal estándar. Recíprocamente, demuestre que si Z tiene distribución normal estándar, y $\sigma > 0$, entonces la variable $X = \mu + \sigma Z$ tiene distribución $\mathrm{N}(\mu, \sigma^2)$.

187. Sea X con distribución $\mathrm{N}(10, 25)$. Calcule
    a) $P(X \geq 10)$.
    b) $P(X < 0)$.
    c) $P(0 < X \leq 10)$.
    d) $P(X \geq 20)$.
    e) $P(-20 < X \leq 10)$.

188. Sea X con distribución $\mathrm{N}(0, 100)$. Calcule
    a) $P(X \leq 10)$.
    b) $P(X > 0)$.
    c) $P(0 < X \leq 40)$.
    d) $P(X \geq 30)$.
    e) $P(-10 < X \leq 10)$.

189. Encuentre x tal que
    a) $\Phi(x) = 0.8666$. b) $1 - \Phi(x) = 0.9154$.

190. Sea X una variable aleatoria con distribución $\mathrm{N}(\mu, \sigma^2)$. Demuestre que la variable $Y = aX + b$, con $a \neq 0$, también tiene una distribución normal. Encuentre los parámetros correspondientes.

191. Sea X una variable aleatoria con distribución $\mathrm{N}(\mu, \sigma^2)$. Demuestre que la variable $Y = -X$ también tiene una distribución normal. Encuentre los parámetros correspondientes.

192. Sea X una variable aleatoria con distribución $\mathrm{N}(0,1)$. Demuestre que X^2 tiene una distribución $\chi^2(1)$. Sugerencia: Para $x>0$, $F_{X^2}(x)= P(X^2 \leq x)= P(-\sqrt{x} \leq X \leq \sqrt{x})$.

193. Sea X una variable aleatoria con distribución normal estándar. Encuentre la función de densidad de $Y = |X|$.

194. Una máquina automática despachadora de refresco en un restaurant está ajustada para llenar vasos de 300 ml en promedio. Debido a cuestiones mecánicas, el llenado de los vasos no es enteramente exacto y hay pequeñas fluctuaciones en el llenado. El fabricante de la máquina sabe que el llenado de los vasos se puede modelar como una v.a. normal con media de 300 ml. y desviación estándar $\sigma = 10$ ml.

    a) ¿Qué fracción de los vasos serán servidos con más de 310 mililitros?

    b) ¿Cuál es la probabilidad de que un vaso sea servido entre 290 y 305 mililitros?

    c) Si el restaurant utiliza vasos de 320 ml. ¿qué porcentaje de ellos se derramarán?

    d) Si los clientes reclaman por vasos servidos con 270 ml. o menos, de mil clientes, ¿cuántos de ellos reclamarán?

Distribución ji cuadrada

195. Sea X con distribución $\chi^2(n)$. Compruebe que la correspondiente función de densidad $f(x)$ efectivamente lo es.

196. Compruebe que la distribución $\chi^2(n)$ se reduce a la distribución $\exp(1/2)$ cuando $n=2$.

197. Sea X con distribución $\chi^2(n)$. Demuestre que $E(X)=n$, y que $\mathrm{Var}(X)=2n$. Compruebe además que $E(X^m) = 2^m\,\Gamma(\frac{n}{2}+m)/\Gamma(\frac{n}{2})$, para $m \geq 0$.

Distribución t

198. Sea X con distribución $t(n)$. Verifique que $f(x)$ es efectivamente una función de densidad.

199. Sea X con distribución $t(n)$. Demuestre que $E(X) = 0$, y $\text{Var}(X) = n/(n-2)$, para $n > 2$.

Vectores aleatorios

200. Sea (X, Y) un vector aleatorio discreto con función de densidad $f(x, y)$ dada por la siguiente tabla

x\y	-1	0
1	.3	.5
2	.05	.15

a) Grafique $f(x, y)$ y demuestre que efectivamente se trata de una función de densidad.

b) Calcule y grafique las densidades marginales $f(x)$ y $f(y)$. Verifique que ambas son efectivamente funciones de densidad.

c) Calcule y grafique las distribuciones marginales $F(x)$ y $F(y)$. Verifique que ambas son efectivamente funciones de distribución.

d) ¿Son X y Y independientes?

201. Sea (X, Y) un vector aleatorio discreto con función de densidad $f(x, y)$ dada por la siguiente tabla

x\y	0	1
1	.2	0
2	.7	.1

a) Grafique $f(x, y)$ y demuestre que efectivamente se trata de una función de densidad.

b) Calcule y grafique las densidades marginales $f(x)$ y $f(y)$. Verifique que ambas son efectivamente funciones de densidad.

c) Calcule y grafique las distribuciones marginales $F(x)$ y $F(y)$. Verifique que ambas son efectivamente funciones de distribución.

d) ¿Son X y Y independientes?

202. Sea (X, Y) un vector aleatorio continuo con distribución uniforme en el cuadrado $(-1, 1) \times (-1, 1)$.

a) Grafique $f(x, y)$ y demuestre que efectivamente se trata de una función de densidad.

b) Calcule y grafique las densidades marginales $f(x)$ y $f(y)$. Verifique que ambas son efectivamente funciones de densidad.

c) Calcule y grafique las distribuciones marginales $F(x)$ y $F(y)$. Verifique que ambas son efectivamente funciones de distribución.

d) ¿Son X y Y independientes?

203. Sea (X,Y) un vector aleatorio continuo con función de densidad $f(x,y)$ dada por la siguiente expresión

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

a) Grafique $f(x,y)$ y demuestre que efectivamente se trata de una función de densidad.

b) Calcule y grafique las densidades marginales $f(x)$ y $f(y)$. Verifique que ambas son efectivamente funciones de densidad.

c) Calcule y grafique las distribuciones marginales $F(x)$ y $F(y)$. Verifique que ambas son efectivamente funciones de distribución.

d) ¿Son X y Y independientes?

Variables y tipos de datos

204. Clasifique cada una de las siguiente variables de acuerdo a si es cualitativa o cuantitativa, y también de acuerdo a su posible escala de medición.

a) Evaluación de un curso usando la siguiente escala: MB=Muy Bien. B=Bien. S=Suficiente. NA=No Acreditado.

b) Número de hijos de una persona.

c) Escolaridad de una persona.

d) Porcentaje de artículos defectuosos producidos en una fábrica.

e) Nivel de satisfacción de una persona al adquirir un producto.

f) Preferencia electoral por alguna de las siguientes propuestas políticas: A,B,C,D,E,F.

g) Uso o no uso de lentes de una persona.

h) Principal país de visita para una persona que viaja al extranjero.

i) Costo económico en un accidente automovilístico.

Estadística descriptiva

205. Calcule la media, moda, mediana, varianza, desviación estándar y rango del siguiente conjunto de datos: $4, 2, 0, 9, 4, 2, -1, 1, -4, 2$.

206. Pregunte a diez personas sus estaturas. Registre todos estos datos y calcule la media, moda, mediana, varianza, desviación estándar y rango.

207. Mediante el uso de una calculadora genere diez números al azar dentro del intervalo unitario $(0, 1)$. Escriba estos datos y calcule la media, moda, mediana, varianza, desviación estándar y rango.

208. Demuestre que $\displaystyle\sum_{i=1}^{n} (x_i - \bar{x}) = 0$.

209. Demuestre que $\displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$.

210. Demuestre que $\displaystyle s^2 = \frac{1}{n-1}\,[\,\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2\,]$.

211. Encuentre el valor de c que minimiza la función $\displaystyle\bar{x}(c) = \sum_{i=1}^{n}(x_i - c)^2$.

212. Calcule la media, varianza, desviación estándar, mediana y moda aproximados del siguiente conjunto de datos agrupados. Elabore además un histograma.

Intervalo de clase	Frecuencia
$10 \leq x < 20$	4
$20 \leq x < 30$	3
$30 \leq x < 40$	6
$40 \leq x < 50$	5
$50 \leq x < 60$	5

213. Calcule la media, varianza, desviación estándar, mediana y moda aproximados del siguiente conjunto de datos agrupados. Elabore además un histogra-

ma.

Intervalo de clase	Frecuencia
$0 \leq x < 5$	12
$5 \leq x < 10$	23
$10 \leq x < 15$	10
$15 \leq x < 20$	14
$25 \leq x < 30$	6
$30 \leq x < 35$	10
$35 \leq x < 40$	5

Muestras aleatorias y estadísticas

214. —

Estimación puntual

215. Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados para el parámetro θ, y sea α una constante. Demuestre que $\hat{\theta} = \alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2$ es también un estimador insesgado para θ,

216. Sea $X_1, \ldots, X_n$ una m.a. de una población con media conocida μ y varianza σ^2 desconocida. Demuestre que la siguiente estadística es un estimador insesgado para σ^2,
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

217. Sea $X_1, \ldots, X_n$ una m.a. de una población con media desconocida y varianza σ^2 desconocida. Demuestre que la siguiente estadística es un estimador insesgado para σ^2,
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

218. Sea $X_1, \ldots, X_n$ una m.a. de una población con media desconocida y varianza finita σ^2 desconocida. Demuestre que la siguiente estadística es un estimador insesgado para σ^2,
$$\hat{\sigma}^2 = \frac{1}{2(n-1)}\sum_{i=1}^{n-1}(X_{i+1} - X_i)^2.$$

219. Sea $X_1, \ldots, X_n$ una m.a. de una población con media μ y varianza σ^2 desconocidas. Recuerde que la varianza muestral S^2 se ha definido como sigue $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Demuestre que S no es un estimador insesgado para σ. Modifique S para que sea insesgado. Sugerencia: $\frac{(n-1)}{\sigma^2}S^2 \sim \chi^2(n-1)$.

Método de momentos

220. Dada una muestra aleatoria de tamaño n de una población uniforme en el intervalo $[0, a]$, use el método de momentos para encontrar un estimador para el parámetro a.

221. Dada una muestra aleatoria de tamaño n de una población Poisson con parámetro desconocido $\lambda > 0$, use el método de momentos para encontrar un estimador del parámetro λ.

222. Dada una muestra aleatoria de tamaño n de una población exponencial con parámetro desconocido $\lambda > 0$, use el método de momentos para encontrar un estimador del parámetro λ.

Método de máxima verosimilitud

223. Dada una muestra aleatoria de tamaño n de una población uniforme en el intervalo $[0, a]$, use el método de máxima verosimilitud para encontrar un estimador para el parámetro a.

224. Dada una muestra aleatoria de tamaño n de una población Poisson con parámetro $\lambda > 0$, use el método de máxima verosimilitud para encontrar un estimador del parámetro λ.

Estimación por intervalos

225. Se sabe que la vida en horas de un foco de 100 watts de cierta marca tiene una distribución aproximada normal con desviación estándar $\sigma = 30$ horas. Se toma una muestra al azar de 50 focos y resultó que la vida media fue de 1550 horas. Construya un intervalo de confianza del 95 % para el verdadero promedio de vida de estos focos.

226. Se realizan 20 pruebas de resistencia de un cierto material obteniéndose los siguientes datos: 2225, 2300, 2217, 2190, 2295, 2285, 2195, 2255, 2232, 2252, 2272, 2231, 2223, 2211, 2219, 2231, 2218, 2262, 2257, 2261. Construya un intervalo de confianza del 98 % para la resistencia media de este material, suponiendo una distribución normal.

Pruebas de hipótesis

227. Se cree que la estatura de cierta población humana sigue una distribución normal con media de 1.70 metros. Realice la prueba de hipótesis $H_0 : \mu = 1.70$ *vs* $H_1 : \mu \neq 1.70$, con el siguiente conjunto de datos: 1.65, 1.75, 1.63, 1.81, 1.74, 1.59, 1.73, 1.66, 1.65, 1.83, 1.77, 1.74, 1.64, 1.69, 1.72, 1.66, 1.55, 1.60, 1.62. Use un nivel de significancia del 10 %, y suponga $\sigma = 0.10$.

228. Las mediciones del número de cigarros fumados al día por un grupo de diez fumadores es el siguiente: $5, 10, 3, 4, 5, 8, 20, 4, 1, 10$. Realice la prueba de hipótesis $H_0 : \mu = 10$ *vs* $H_1 : \mu < 10$, suponiendo que los datos provienen de una muestra tomada al azar de una población normal con $\sigma = 1.2$. Use un nivel de significancia del 5 %.

229. Considere cualquiera de las tres pruebas de hipótesis para la media μ de una población normal con σ^2 conocida. Demuestre que en cualquier caso la probabilidad de cometer el error tipo II tiende a cero cuando el tamaño de la muestra n tiende a infinito.

Apéndice B

Soluciones

Esta sección contiene algunas sugerencias de solución a los ejercicios planteados. Para algunos ejercicios es necesario ser más explícito al dar una solución o al justificar una respuesta, considere por tanto que este material contiene simplemente ideas para generar una solución completa y bien escrita. La mayoría de las gráficas han sido omitidas. Recuerde además que los métodos empleados o sugeridos para llegar a una solución no son necesariamente únicos.

Espacio muestral y eventos

1. *a*) $\Omega = \{AAA, AAS, ASA, ASS, SAA, SAS, SSA, SSS\}$.
   b) $\Omega = \{AAA, AAS, ASS, SSS\}$.
   c) $\Omega = [0, 1]$.
   d) $\Omega = \{0, 1, 2, \ldots\}$.
   e) Considere que las bolas tienen las etiquetas 1 y 2. Considere el arreglo (C_1, C_2, C_3, C_4) en donde cada coordenada C_i corresponde al conjunto de bolas que han sido depositadas en la celda i. Entonces $\Omega = \{(\{1, 2\}, \emptyset, \emptyset, \emptyset), (\emptyset, \{1, 2\}, \emptyset, \emptyset), \ldots, (\{1\}, \{2\}, \emptyset, \emptyset), \ldots\}$. El número total de elementos en Ω es 16. ¿Puede usted escribir todos los arreglos?
   f) $\Omega = \{(2, 0, 0, 0), (0, 2, 0, 0), (0, 0, 2, 0), (0, 0, 0, 2), (1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1), (0, 1, 1, 0), (0, 1, 0, 1), (0, 0, 1, 1)\}$.
   g) $\Omega = \{(0, 0), (1, 0), (0, 1), (1, 1), \ldots\}$.

2. Si nos interesa observar el número de lanzamientos hasta obtener un 6 se puede tomar $\Omega = \{1, 2, \ldots\}$. Si nos interesa registrar con detalle los resultados, $\Omega = \{6, N6, NN6, NNN6, \ldots\}$, en donde cada N puede ser cualquiera de los números $1, 2, 3, 4$ ó 5.

3. a) Número de hijos de una persona escogida al azar. b) Tiempo que una persona tiene que esperar para la llegada de un autobús. c) Este es el espacio muestral más frecuente. El número 1 puede representar ganar en un juego de lotería mientras que el número 0 corresponde a perder en el juego. Este es uno de los muchos ejemplos que pueden construirse para este espacio.

Operaciones con conjuntos

4. a) Sea $x \in A \cap (B \cap C)$. Entonces $x \in A$ y $x \in B \cap C$. La última afirmación puede descomponerse en: $x \in B$ ó $x \in C$, esto incluye el caso de pertenencia a ambos conjuntos. Si es el primer caso, es decir $x \in B$, entonces se obtiene que $x \in A \cap B$, y por lo tanto x es un elemento del lado derecho de la igualdad. La misma conclusión se obtiene si $x \in C$. Esto demuestra la contención $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$. De manera análoga se prueba la contención contraria. b) Análogo al inciso anterior.

5. a) Demostraremos la contención $(A \cup B)^c \subseteq A^c \cap B^c$. Sea $x \in (A \cup B)^c$. Entonces $x \notin (A \cup B)$, esto es, $x \notin A$ y $x \notin B$. Por lo tanto $x \in A^c$ y $x \in B^c$. Es decir, $x \in A^c \cap B^c$. De manera análoga se prueba la contención contraria y de esa forma se demuestra la igualdad. b) Pruebe ambas contenciones como en el inciso anterior.

6. Los enunciados son: $a)(A \cup B \cup C)^c = A^c \cap B^c \cap C^c$ y $b)(A \cap B \cap C)^c = A^c \cup B^c \cup C^c$. Para el primer inciso se tiene que $(A \cup B \cup C)^c = ((A \cup B) \cup C)^c = (A \cup B)^c \cap C^c = A^c \cap B^c \cap C^c$. De manera análoga se demuestra al segundo inciso.

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

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

9. Análogo al anterior.

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

11. Análogo al anterior.

12. a) $A - (A \cap B) = A \cap (A \cap B)^c = A \cap (A^c \cup B^c) = (A \cap A^c) \cup (A \cap B^c) = \emptyset \cup (A \cap B^c) = A \cap B^c = A - B$. b) $(A \cup B) - B = (A \cup B) \cap B^c = (A \cap B^c) \cup (B \cap B^c) = (A \cap B^c) \cup \emptyset = A \cap B^c = A - B$.

13. a) $(A \cap B) - (A \cap C) = (A \cap B) \cap (A \cap C)^c = (A \cap B) \cap (A^c \cup C^c) = (A \cap B \cap A^c) \cup (A \cap B \cap C^c) = \emptyset \cup (A \cap B \cap C^c) = A \cap (B \cap C^c) = A \cap (B - C)$. b) $(A - C) \cap (B - C) = (A \cap C^c) \cap (B \cap C^c) = A \cap (B \cap C^c) = A \cap (B - C)$.

14. $(A \cup B) - (A \cap B) = (A \cup B) \cap (A \cap B)^c = (A \cup B) \cap (A^c \cup B^c) = [(A \cup B) \cap A^c] \cup [(A \cup B) \cap B^c] = [B \cap A^c] \cup [A \cap B^c] = (B - A) \cup (A - B)$.

15. Compruebe que ambos lados de la igualdad del inciso (a) corresponden al evento que se muestra en la Figura B.1. El resto de los incisos de obtiene fácilmente de la definición.

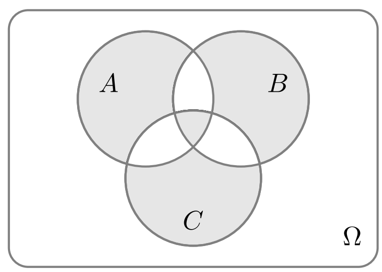


Figura B.1: $A \triangle B \triangle C$.

16. Gráficamente los conjuntos A y B son como se muestran en la Figura B.2. Se tiene entonces que $A = [-2, 6]$, $B = (-\infty, -1) \cup (3, \infty)$, $A^c = (-\infty, -2) \cup (6, \infty)$, $B^c = [-1, 3]$, $A \cap B = [-2, -1) \cup (3, 6]$, $A \cup B = \mathbb{R}$, $A - B = [-1, 3]$, $B - A = (-\infty, -2) \cup (6, \infty)$, $A \triangle B = (-\infty, -2) \cup [-1, 3] \cup (6, \infty)$.

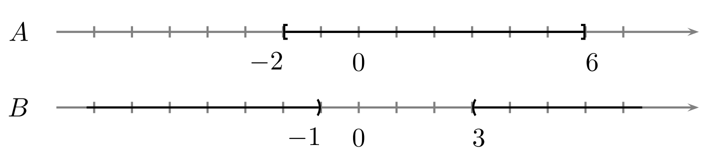


Figura B.2:

17. Gráficamente los conjuntos A y B son los que se muestran en la Figura B.3. Por lo tanto $A = [-3, 1]$, $B = (-\infty, -\sqrt{2}) \cup (\sqrt{2}, \infty)$, $A^c = (-\infty, -3) \cup (1, \infty)$, $B^c = [-\sqrt{2}, \sqrt{2}]$, $A \cap B = [-3, -\sqrt{2})$, $A \cup B = (-\infty, 1) \cup (\sqrt{2}, \infty)$, $A - B = [-\sqrt{2}, 1]$, $B - A = (-\infty, -3) \cup (\sqrt{2}, \infty)$, $A \triangle B = (-\infty, -3) \cup [-\sqrt{2}, 1] \cup (\sqrt{2}, \infty)$.

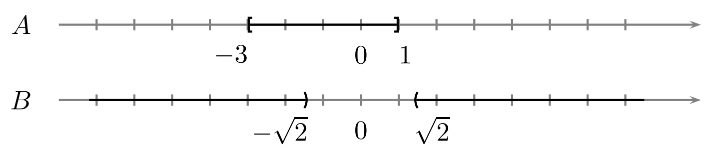


Figura B.3:

18. Los conjuntos A y B son los que se muestra en la Figura B.4 (a). Allí pueden identificarse gráficamente las operaciones que se solicitan.

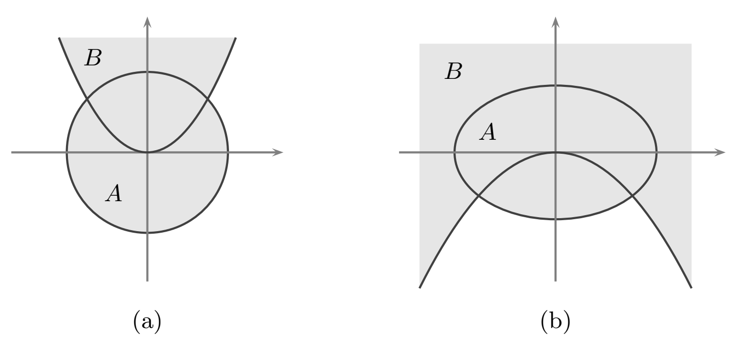


Figura B.4:

19. Los conjuntos A y B son los que se muestran en la Figura B.4 (b). Allí pueden identificarse gráficamente las operaciones que se solicitan.

20. El máximo es 50 % y el mínimo es 0 %.

21. El máximo es 70 % y el mínimo es 40 %.

Conjuntos ajenos

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

23. No son ajenos, tienen un único punto en común que es $(0, 1)$.

24. $A_1 \cap B \subseteq A \cap B = \emptyset$.

25. $n(n-1)/2$.

Conjunto potencia

26. $2^{\emptyset} = \{\emptyset\}$.

27. Sea $A_1 \in 2^A$, entonces $A_1 \subseteq A \subseteq B$. Por lo tanto $A_1 \subseteq B$ y en consecuencia $A_1 \in 2^B$.

Producto Cartesiano

28. $A \times B \times C = \{ (a_1, b_1, c_1), (a_1, b_1, c_2), (a_1, b_2, c_1), (a_1, b_2, c_2), (a_1, b_3, c_1), (a_1, b_3, c_2), (a_2, b_1, c_1), (a_2, b_1, c_2), (a_2, b_2, c_1), (a_2, b_2, c_2), (a_2, b_3, c_1), (a_2, b_3, c_2)\}$.

29. $A \times A = \{ (a, a), (a, b), (a, c), (b, a), (b, b), (b, c), (c, a), (c, b), (c, c)\}$.

Probabilidad

30. Claramente $P(A) = \#A/\#\Omega \geq 0$, y $P(\Omega) = \#\Omega/\#\Omega = 1$. Además, cuando A y B son ajenos, $\#(A \cup B) = \#A + \#B$, y de allí se sigue la tercera propiedad.

31. Para cualquier valor natural de n, se cumple que $n(A)/n \geq 0$, y $n(\Omega)/n = n/n = 1$. Cuando A y B son ajenos, $n(A \cup B) = n(A) + n(B)$. Ahora tome límite cuando n tiende a infinito y obtenga el resultado requerido.

32. a) $P(\Omega) = \alpha P_1(\Omega) + (1-\alpha)P_2(\Omega) = \alpha + (1-\alpha) = 1$. b) $P(A) = \alpha P_1(A) + (1-\alpha)P_2(A) \geq 0$. c) Cuando A y B son disjuntos, $P(A \cup B) = \alpha P_1(A \cup B) + (1-\alpha)P_2(A \cup B) = \alpha(P_1(A) + P_1(B)) + (1-\alpha)(P_2(A) + P_2(B)) = [\alpha P_1(A) + (1-\alpha)P_2(A)] + [\alpha P_1(B) + (1-\alpha)P_2(B)] = P(A) + P(B)$.

33. $P(\emptyset) = P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset)$. Ahora cancele uno de los términos.

34. $B = (B \cap A) \cup (B \cap A^c)$, siendo esta unión ajena. Aplicando probabilidad, $P(B) = P(B \cap A) + P(B \cap A^c)$. El segundo sumando es $P(B - A)$.

35. La primera desigualdad es evidente. Como $A \cap B \subseteq A$, se tiene que $P(A \cap B) \leq P(A)$. Análogamente como $A \subseteq A \cup B$, $P(A) \leq P(A \cup B)$. La siguiente desigualdad es consecuencia de la fórmula $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Finalmente la última desigualdad se obtiene de $P(A) \leq P(\Omega) = 1$.

36. El resultado se obtiene de $1 \geq P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

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

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

39. a) $P(A \cup B) = 0.5$. b) $P(A^c) = 0.7$. c) $P(A^c \cap B) = 0.2$. d) $P(A \cap B^c) = 0.3$. e) $P(A \triangle B) = 0.5$.

40. a) Verdadero, pues $A \cap B \subseteq A$. b) Falso, considere por ejemplo el experimento de lanzar una moneda equilibrada. Ambos resultados tienen la misma probabilidad pero evidentemente no son iguales. c) Falso, en el experimento de lanzar un dado equilibrado, considere por ejemplo los eventos $A = \{1\}$ y $B = \{2, 3\}$. Claramente $P(A) \leq P(B)$ pero A no está contenido en B.

41. a) Verdadero, pues $A \subseteq A \cup B$. b) Falso, A y B pueden ser ajenos. c) Verdadero, $P(A \cup B) = P(A) + P(B) - P(A \cap B) > 1/2 + 1/2 - P(A \cap B) = 1 - P(A \cap B) \geq 0$. d) Falso, A puede ser Ω.

42. a) Falso. En el experimento de lanzar una moneda equilibrada, sea A el resultado de obtener una de las caras y B obtener la otra cara. Entonces $P(B - A) = P(B) = 1/2$, mientras que $P(B) - P(A) = 1/2 - 1/2 = 0$. El resultado es cierto bajo la condición $A \subseteq B$. b) Falso, excepto cuando A y B son independientes. Considerando el ejemplo del inciso anterior, $P(A \cap B) = 0$, mientras que $P(A)P(B) = 1/2 \cdot 1/2 = 1/4$. c) Verdadero. $P(A^c) = 1 - P(A) < 1 - 1/2 = 1/2$.

43. $P(A) = 2/3$.

Análisis combinatorio

44. $6! = 720$.

45. El total de casos posibles es $5! = 120$. Los casos favorables son dos: 12345 y 54321. Por lo tanto la probabilidad buscada es $2/120 = 1/60$.

46. $\binom{n}{2} - n = n(n-1)/2$.

47. $15 \cdot 14 \cdot 13 = 2730$.

48. Un entero positivo de a lo sumo cinco dígitos que es divisible por 2 es de la forma $xxxxy$ en donde x puede ser cualquiera de $0, 1, \ldots, 9$ y y puede ser $0, 2, 4, 6, 8$. Por lo tanto el total es $10 \cdot 10 \cdot 10 \cdot 10 \cdot 5 = 50000$, excepto que debemos omitir el entero 00000. Si empiezan con el número 1, entonces son de la forma $1xxxy$, y entonces la respuesta es $10 \cdot 10 \cdot 10 \cdot 5 = 5000$ y esta vez no hay necesidad de omitir ninguno de ellos.

49. $\left(\begin{array}{c} 5 \\ 3 \end{array} \right) = 5!/3!2! = 10$.

50. Simplifique el lado derecho.

51. Simplifique el lado derecho.

52. Simplifique el lado derecho.

53. Simplifique el lado derecho.

54. $\binom{50}{k}\binom{200}{20-k}/\binom{250}{20}$.

55. Considerando inicial y artificialmente que los siete dígitos son todos distintos, la respuesta preliminar es $7! = 5040$. Como los cuatro dígitos 1 son idénticos, cualquier permutación entre ellos produce el mismo número y por lo tanto debemos dividir el resultado preliminar por 4! Análogamente con los tres dígitos 0. La respuesta final es $7!/4!\,3! = 5040/(24 \cdot 6) = 35$.

56. Razonando como en el ejercicio anterior, la respuesta es $4!/(2!\,1!\,1!) = 6$.

57. Considerando que cada año tiene 365 dias, la respuesta es $1 - 365 \cdot 364 \cdots (365 - n + 1)/365^n$.

58. El total es 4536. No es cierto que la mitad de ellos sean pares y la otra mitad impares. Existen 2296 pares y 2240 impares.

59. —

60. —

61. —

62. —

63. —

64. —

65. —

66. —

67. —

68. —

69. Se pueden obtener un total de 2^n subconjuntos distintos de un conjunto de cardinalidad n. Los 16 subconjuntos de $\{a,b,c,d\}$ son: $\emptyset$, $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$, $\{a,b\}$, $\{a,c\}$, $\{a,d\}$, $\{b,c\}$, $\{b,d\}$, $\{c,d\}$, $\{a,b,c\}$, $\{b,c,d\}$, $\{a,c,d\}$, $\{a,b,d\}$, $\{a,b,c,d\}$.

70. 2^n. El conjunto E corresponde al conjunto de los n ejes coordenados, o uniones de éstos.

71. —

72. —

73. 2^6.

74. $n!\,(n+1)!$

75. $(n!)^2$.

76. $7!/(3!\,2!)$.

77. $11!/(5!\,4!\,2!)$.

78. Considere que se tienen k casillas numeradas $1,2,\ldots,k$ en donde se desean colocar n bolas sin ninguna restricción en el número de bolas que caben en cada casilla. De esta forma, el número de bolas en la casilla 1 es el valor de x_1, lo mismo para las casillas $2,3,\ldots,k$. Las respuestas son entonces
a) $\binom{n+k-1}{n}$. b) $\sum_{m=0}^{n}\binom{m+k-1}{m}$. c) Infinito.

79. Considere que los números $1, 2, \ldots, n$ representan casillas que se encuentran ordenadas de manera natural. El problema puede entonces traducirse en colocar k bolas en este arreglo de casillas en donde cada casilla admite a lo sumo una bola. La respuesta es entonces $\binom{n}{k}$.

80. $(a+b+c)^3 = a^3 + b^3 + c^3 + 3ab^2 + 3ac^2 + 3a^2b + 3a^2c + 3b^2c + 3bc^2 + 6abc$.

Probabilidad condicional e independencia

81. a) Como $B = (A^c \cap B) \cup (A \cap B)$, se tiene que $P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(B)P(A^c)$. b) Análogo al primer inciso. c) $P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A \cap B) = 1 - P(A) - P(B) + P(A)P(B) = (1 - P(A))(1 - P(B)) = P(A^c)P(B^c)$.

82. a) $P(A \cap B) = 0.05$ b) $P(A^c \cap B) = 0.45$ c) $P(A \cap B^c) = 0.05$ d) $P(A \cup B) = 0.55$ e) $P(A^c \cup B) = 0.95$ f) $P(A \cup B^c) = 0.55$.

83. a) $P(\Omega|B) = P(B)/P(B) = 1$ b) $P(A|B) \geq 0$ c) Sean A_1 y A_2 ajenos. Entonces $A_1 \cap B$ y $A_2 \cap B$ también son ajenos. Por lo tanto $P(A_1 \cup A_2|B) = P((A_1 \cup A_2) \cap B)/P(B) = P((A_1 \cap B) \cup (A_2 \cap B))/P(B) = P(A_1 \cap B)/P(B) + P(A_2 \cap B)/P(B) = P(A_1|B) + P(A_2|B)$.

84. $P(A \cap B \cap C) = P(C|A \cap B)P(A|B)P(B) = ppp = p^3$.

85. a) Falso. Con $A = \emptyset$ se obtiene $P(A|B) = 0$. b) Falso. Tómese $A = \Omega$ y B tal que $0 < P(B) < 1$. Entonces $P(A|B) = 1$ mientras que $P(B|A) = P(B) < 1$. c) Falso. Tómese $A = B$ con $0 < P(A) < 1$. Entonces $P(A|B) = 1$ mientras que $P(A) < 1$.

86. a) Verdadero y ello es consecuencia del hecho de que $P(\cdot|B)$ es una medida de probabilidad. Alternativamente $P(A|B) + P(A^c|B) = P(A \cap B)/P(B) + P(A^c \cap B)/P(B) = P((A \cap B) \cup (A^c \cap B))/P(B) = P((A \cup A^c) \cap B)/P(B) = P(B)/P(B) = 1$. b) Falso. Tómese $A = \Omega$. El lado izquierdo es 2 mientras que el lado izquierdo es 1. c) Cierto. $P(A|A \cap B) = P(A \cap B)/P(A \cap B) = 1$. El cálculo es idéntico para $P(B|A \cap B)$.

87. Todos los incisos son ciertos. Solo es cuestión de escribir la definición de cada probabilidad condicional.

88. a) Falso. Tómese por ejemplo $A = B_1 \cup B_2$. Entonces el lado izquierdo es uno mientras que el lado derecho es dos. b) Cierto. $P(A_1 \cup A_2|B) = P((A_1 \cup$

$A_2)\cap B)/P(B) = P(A_1\cap B)/P(B)+P(A_2\cap B)/P(B) = P(A_1|B)+P(A_2|B)$.
c) Falso. Tómese $B_1 = B_2$. El lado derecho es el cuadrado del lado izquierdo.
d) Falso. Tómese $A_1 = A_2$. El lado derecho es el cuadrado del lado izquierdo.

89. a) $P(\emptyset\cap\emptyset) = P(\emptyset) = 0 = P(\emptyset)P(\emptyset)$. b) $P(\Omega\cap\Omega) = P(\Omega) = 1 = P(\Omega)P(\Omega)$.

90. a) Falso. Tómese $B = \Omega$. Entonces A y B son independientes pero no son ajenos. b) Falso. Considere el experimento de lanzar una moneda honesta, y sean A y B los eventos de que caiga una cara y la otra respectivamente. Entonces claramente A y B son ajenos pero no son independientes.

91. $P(A \cap \emptyset) = P(\emptyset) = 0 = P(A)P(\emptyset)$.

92. $P(A \cap \Omega) = P(A) = P(A)P(\Omega)$.

93. $P(A \cup B) = 1 - P(A^c \cap B^c) = 1 - P(A^c)P(B^c)$ pues A^c y B^c también son independientes. Alternativamente desarrolle $P(A \cup B)$ y factorice adecuadamente.

94. a) Cierto pues la definición es simétrica respecto de los eventos A y B. b) Falso en general. Tómese A tal que $0 < P(A) < 1$. Entonces $P(A\cap A) = P(A)$ mientras que $P(A)P(A) < P(A)$. Cuando $P(A)$ es cero o uno la propiedad es válida. c) Falso. Tómese $\Omega = \{1, 2, 3, 4\}$ equiprobable con $A = \{1, 2\}$, $B = \{2, 3\}$ y $C = \{3, 4\}$. Entonces se cumple $P(A\cap B) = 1/4 = P(A)P(B)$ y $P(B\cap C) = 1/4 = P(B)P(C)$, pero $P(A\cap C) = 0$ distinto a $P(A)P(C) = 1/4$.

95. La probabilidad de que no ocurra ninguno de los eventos es $P(A^c \cap B^c)$, que por la independencia es igual a $P(A^c)P(B^c) = (1 - p_1)(1 - p_2)$.

96. Como A y B son independientes y ajenos, $0 = P(A \cap B) = P(A)P(B)$. Para que este producto se anule forzosamente alguno de los factores tiene que ser cero.

97. a) $\emptyset = (0, 0), (0, 1/2)$. b) $(0, 1/2), (1/4, 3/4)$. c) $(0, 1/2), (1/2, 1)$.
d) $(0, 1/2), (0, 1/2)$.

98. El total de subconjuntos de cardinal $1, 2, \ldots, n$ es respectivamente $\binom{n}{1}$, $\binom{n}{2}$, $\ldots, \binom{n}{n}$. De modo que la respuesta es $\binom{n}{1} + \cdots + \binom{n}{n} = 2^n - n - 1$.

Teorema de probabilidad total

99. Se puede usar el método de inducción sobre n. Claramente $P(R_1) = r/(r+b)$. Suponga entonces válida la igualdad $P(R_n) = r/(r+b)$. Se condiciona la probabilidad del evento R_{n+1} dado el resultado de la primera extracción, y en tal situación puede considerarse que la configuración inicial de la urna cambia, de modo que el evento R_{n+1} se refiere ahora a la n-ésima extracción y allí es donde se usa la hipótesis de inducción. Entonces por el teorema de probabilidad total,

$$\begin{aligned} P(R_{n+1}) &= P(R_{n+1}|R_1)P(R_1) + P(R_{n+1}|R_1^c)P(R_1^c) \\ &= \frac{r+c}{r+b+c} \times \frac{r}{r+b} + \frac{r}{r+b+c} \times \frac{b}{r+c} \\ &= \frac{r}{r+b}. \end{aligned}$$

100. Sean N_1, N_2 y N_3 los eventos correspondientes a escoger los números 1, 2 y 3 respectivamente. Sea A el evento "obtener 5 al lanzar el dado". Entonces $P(A) = P(A|N_1)P(N_1) + P(A|N_2)P(N_2) + P(A|N_3)P(N_3) = 11/108$.

Teorema de Bayes

101. Sea A el evento cuando todas las bolas son del mismo color, y N cuando en particular son todas negras. Entonces $P(N|A) = P(A|N)P(N)/P(A) = \binom{n}{k}/[\binom{n}{k} + \binom{b}{k}]$.

102. Sea R_0 el evento "se recibe un cero", y E_0 el evento "se envía un cero". Análogamente se definen los eventos R_1 y E_1. Entonces

    a) $P(R_0) = P(R_0|E_0)P(E_0) + P(R_0|E_1)P(E_1) = 0.8 \cdot 0.4 + 0.1 \cdot 0.6 = 0.38$.
    b) $P(R_1) = P(R_1|E_0)P(E_0) + P(R_1|E_1)P(E_1) = 0.2 \cdot 0.4 + 0.9 \cdot 0.6 = 0.62$.
    c) $P(E_0|R_0) = P(R_0|E_0)P(E_0)/P(R_0) = 0.8 \cdot 0.4/0.38 = 0.84$.
    d) $P(E_1|R_1) = P(R_1|E_1)P(E_1)/P(R_1) = 0.9 \cdot 0.6/0.62 = 0.87$.

Variables aleatorias

103. a) Discreta, los valores pueden ser $0, 1, 2, \ldots$ si la escala de medición son años. b) Discreta con posibles valores $0, 1, 2, \ldots$ c) Puede ser continua y

considerarse teóricamente el intervalo $[0, \infty)$. d) Discreta con posibles valores $0, 1, 2, \ldots$ si sólo se consideran unidades monetarias.

104. a) Continua con valores en $(0, 1)$. b) Discreta con valores en el conjunto $\{0, 1, 2, \ldots, 8, 9\}$. c) Continua con valores en $(0, 1)$. d) Continua con valores en $(0, 1/2)$.

105. El rango de X es el conjunto $\{-4, -2, 0, 2, 4, 6\}$. $P(X = 0) = 1/6$, $P(X \in \{2, 3\}) = 1/6$, $P(X \geq 0) = 2/3$, $P(X < 0) = 1/3$, $P(X^2 = 1) = 0$, $P(2X - 4 = 0) = 1/6$ y $P(X^2 = 4) = 1/3$.

106. $P(X \geq 0) = 1/2$, $P(X < 0) = 1/2$, $P(Y > X) = 1/2$, $P(X + Y \leq 1) = 3/4 + 2/\pi$, $P(Z < 1/2) = 1/4$ y $P(1/3 < Z < 1/2) = 5/36$.

Funciones de densidad y de distribución

107. a) $c = 1/55$, $P(X \in \{2, 3, 4\}) = 9/55$, $P(X < 3) = 3/55$. b) $c = 1/385$, $P(X \in \{2, 3, 4\}) = 9/385$, $P(X < 3) = 3/385$.

108. $c = 1/2\pi$, $P(X \geq \pi) = (\pi - 2)/2\pi$, $P(X \in [-\pi, 2\pi]) = 1$.

109. $c = 1/2$, $P(X \in (1, \infty)) = 1/2e$.

110. a) y b) La constante c no puede ser cero. Tampoco puede ser positiva o negativa pues en ambos casos la función de probabilidad toma valores negativos, lo cual no es permitido.

111. $P(X \geq 0) = 0.8$, $P(X < 0) = 0$, $P(X^2 = 1) = 0.7$.

112. La función de probabilidad de X^2 es

x	0	1	4	9	25
$f(x)$	0.4	0.15	0.2	0.15	0.1

La de $\lvert X\rvert$ es

x	0	1	2	3	5
$f(x)$	0.4	0.15	0.2	0.15	0.1

Y la de $2X - 5$ es

x	-9	-7	-5	-1	1	5
$f(x)$	0.1	0.15	0.4	0.1	0.15	0.1

113. El valor de c es 0.8. La función de probabilidad de X^2 es

x	0	4
$f(x)$	0.8	0.2

114. a) $k = 2$. b) $F(x) = (x + 2)/10$ para $x \in [-2, 8]$, cero antes, uno después. c) $P(-1 \leq X \leq 3) = 2/5$, $P(X \geq 2) = 3/5$, $P(X \leq 0) = 1/5$. d) $m = 5/2$.

115. Se trata de una variable aleatoria discreta cuya función de probabilidad es $f(1) = 1/3$ y $f(2) = 2/3$. Por lo tanto $P(X = 2) = 2/3$, y $P(1 < X < 2) = 0$.

116. La función de densidad es $f(x) = 1/(2\sqrt{x})$, para $x \in (0, 1)$. $P(X = 1/2) = 0$ y $P(X > 1/2) = F(1) - F(1/2) = 1 - 1/\sqrt{2}$.

117. a) $\Omega = \{(1,2),(1,3),(1,4),(2,1),(2,3),(2,4),(3,1),(3,2),(3,4),(4,1),(4,2)(4,3)\}$. b) La función de probabilidad de X es

x	3	4	5	6	7
$f(x)$	2/12	2/12	4/12	2/12	2/12

d) $P(X \geq 6) = 1/3$, $P(3 < X \leq 5) = 1/2$, $P(X = 6) = 1/6$.

118. Es función de densidad pues es no negativa y es tal que $f(0)+f(1)+f(2) = 1$.

119. Es función de densidad pues es no negativa y por el teorema del binomio se cumple que $\sum_{x=0}^{4} f(x) = (3/4+1/4)^4 = 1$. Esta es la función de probabilidad de la distribución $\text{bin}(n, p)$ con $n = 4$ y $p = 3/4$.

120. Ninguna de las funciones es de densidad. La primera no integra uno, y para la segunda se tiene que $f(3/2) < 0$.

121. La función de probabilidad es $f(x) = (1/2)^{x+1}$, para $x = 0, 1, 2, \ldots$ Por lo tanto $P(0 \leq X < 10) = F(9) = 1 - (1/2)^{10}$.

122. $c = 3$. $F(x) = x^2/9$ para $x \in [0, 3]$, cero antes, uno después.

Esperanza, varianza, momentos

123. Puede considerarse la variable aleatoria constante $X = a$. Alternativamente puede tomarse a X como aquella variable aleatoria que toma los valores $a - 1$ y $a + 1$ con idéntica probabilidad 1/2. ¿Puede usted construir una variable aleatoria continua con la propiedad indicada?

124. a) $E(X) = 7/6$. b) $E(X) = 0$.

125. a) $E(x) = 1$. b) $E(X) = 4/3$.

126. La función es no negativa. Usando fracciones parciales se comprueba la identidad
$$\frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{x+1}.$$
Entonces la suma $\sum_{x=1}^{\infty} f(x)$ es telescópica y vale uno. Por otro lado la esperanza no existe pues la suma $\sum_{x=1}^{\infty} xf(x) = \sum_{x=1}^{\infty} \frac{1}{x+1}$ es divergente.

127. Para demostrar que esta función es de densidad es necesario recordar que la derivada de la función $h(x) = \arctan x$ es $h'(x) = 1/(1+x^2)$. Por lo tanto $\int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)}\, dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{d}{dx} \arctan x\, dx = \frac{1}{\pi} \arctan x|_{-\infty}^{\infty} = \frac{1}{\pi}(\pi/2 - (-\pi/2)) = 1$. La esperanza no existe por que la integral resultante no es absolutamente convergente: $\int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)}\, dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2}\, dx \geq \frac{2}{\pi} \int_1^{\infty} \frac{x}{1+x^2}\, dx \geq \frac{2}{\pi} \int_1^{\infty} \frac{x}{x^2+x^2}\, dx = \frac{1}{\pi} \int_1^{\infty} \frac{1}{x}\, dx = \infty$.

128. a) Cierto, la constante cero cumple tal condición. b) Falso, las siguientes dos distribuciones son distintas y ambas tienen esperanza cero.

x	-1	0	1
$f_1(x)$	1/2	0	1/2
$f_2(x)$	1/3	1/3	1/3

c) Falso, si X es una variable aleatoria con esperanza finita y positiva, entonces $-X$ tiene esperanza negativa. d) Cierto, por ejemplo todas las constantes son variables aleatorias con varianza cero. e) Cierto, pues la varianza es la esperanza de una variable aleatoria no negativa. f) Falso, dos variables aleatorias constantes distintas tienen varianza cero.

129. Tanto la esperanza como la varianza son constantes. Los resultados se siguen del hecho de que la esperanza de una constante es la constante misma y la varianza de una constante es cero.

130. a) $E(X) = c \times P(X = c) = c \times 1 = c$. b) $E(X^n) = c^n \times P(X = c) = c^n \times 1 = c^n$. c) $\text{Var}(X) = (c - E(X))^2 \times P(X = c) = (c - c)^2 \times 1 = 0$.

131. $E(X) = 3$ y $\text{Var}(X) = 24/9$.

132. Esta es la distribución geo(p) con $p = 1/2$. Usaremos resultados de sumas geométricas. $E(X) = \sum_{x=0}^{\infty} x(1/2)^{x+1} = \sum_{x=1}^{\infty} \sum_{y=1}^{x} (1/2)^{x+1}$. Intercambiando el orden de las sumas se encuentra la expresión $\sum_{y=1}^{\infty} \sum_{x=y}^{\infty} (1/2)^{x+1}$

$= \sum_{y=1}^{\infty}(1/2)^y = 1$. Para la varianza encontraremos primero el segundo momento, y para ello usaremos la identidad $x^2 = x(x+1) - x$. $E(X^2) = \sum_{x=1}^{\infty} x^2(1/2)^{x+1} = \sum_{x=0}^{\infty} x(x+1)(1/2)^{x+1} - \sum_{x=0}^{\infty} x(1/2)^{x+1}$. La segunda suma acaba de ser calculada y vale uno. Para la primera suma escriba $x(x+1)$ como $2\sum_{y=1}^{x} y$. Intercambiando ahora el orden de las sumas se llega a $E(X^2) = 2\sum_{y=1}^{\infty} y \sum_{x=y}^{\infty}(1/2)^{x+1} - 1 = 2\sum_{y=1}^{\infty} y(1/2)^y - 1 = 2^2\sum_{y=1}^{\infty} y(1/2)^{y+1} - 1 = 4 - 1 = 3$. Por lo tanto la varianza es $\mathrm{Var}(X) = E(X^2) - E^2(X) = 3 - 1 = 2$.

133. Ambos incisos son ciertos pues tanto $E(X)$ como $\mathrm{Var}(X)$ son constantes.

134. La función es de densidad pues es no negativa y la integral $\int_{-\infty}^{\infty} \frac{1}{2}e^{-|x|}\,dx$ puede ser escrita como $2\int_0^{\infty} \frac{1}{2}e^{-x}\,dx = \int_0^{\infty} e^{-x}\,dx = 1$.
a) La esperanza de X es $\int_{-\infty}^{\infty} x\frac{1}{2}e^{-|x|}\,dx$. Separando la parte positiva y la negativa se obtiene $\int_0^{\infty} x\frac{1}{2}e^{-x}\,dx + \int_{-\infty}^{0} x\frac{1}{2}e^{x}\,dx$. Mediante el cambio de variable $y = -x$ se comprueba que la segunda integral es el negativo de la primera, se obtiene que la esperanza es cero.
b) El segundo momento es $\int_{-\infty}^{\infty} x^2\frac{1}{2}e^{-|x|}\,dx$. Nuevamente separando la parte positiva de la negativa se obtiene $\int_0^{\infty} x^2\frac{1}{2}e^{-x}\,dx + \int_{-\infty}^{0} x^2\frac{1}{2}e^{x}\,dx$. Ahora al hacer el cambio de variable $y = -x$ en la segunda integral se obtiene que es idéntica a la primera integral. Por lo tanto se obtiene el doble de la primera, es decir, $\int_0^{\infty} x^2 e^{-x}\,dx$. Usando integración por partes puede comprobarse que esta integral vale 2.
c) Por los cálculos anteriores $\mathrm{Var}(X) = E(X^2) - E^2(X) = 2$.
d) El n-ésimo momento es $\int_{-\infty}^{\infty} x^n\frac{1}{2}e^{-|x|}\,dx$. Nuevamente separando la parte positiva de la negativa se obtiene $\int_0^{\infty} x^n\frac{1}{2}e^{-x}\,dx + \int_{-\infty}^{0} x^n\frac{1}{2}e^{x}\,dx$. Al hacer el cambio de variable $y = -x$ en la segunda integral se obtiene que es $(-1)^{n+2}\int_0^{\infty} x^n\frac{1}{2}e^{-x}\,dx$. Cuando n es impar se trata del negativo de la primera integral, de modo que la suma es cero. Cuando n es par, es idéntica a la primera integral, y entonces el resultado es $\int_0^{\infty} x^n e^{-x}\,dx$. Aplicando repetidas veces el método de integración por partes se comprueba que esta integral vale $n!$

135. a) Cierto pues la esperanza es lineal. b) Falso pues en general $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$. c) Falso. El lado izquierdo es $\mathrm{Var}(X)$ y el lado derecho es siempre cero.

136. La tercera igualdad no es correcta pues la varianza en general no es lineal.

Distribución uniforme discreta

137. *a*) $E(X) = \sum_{i=1}^{n} i \, \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n} \, \frac{n(n+1)}{2} = (n+1)/2.$

b) $E(X^2) = \sum_{i=1}^{n} i^2 \, \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} i^2 = \frac{1}{n} \, \frac{n(n+1)(2n+1)}{6} = (n+1)(2n+1)/6.$

c) $\mathrm{Var}(X) = E(X^2) - E^2(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = (n^2-1)/12.$

138. 45/100.

Distribución Bernoulli

139. Es muy sencillo verificar que $f(x) \geq 0$, y $f(0) + f(1) = 1$. Además
a) $E(X) = 0\,(1-p) + 1\,p = p$.
b) $E(X^n) = 0^n\,(1-p) + 1^n\,p = p$.
c) $\mathrm{Var}(X) = E(X^2) - E^2(X) = p - p^2 = p(1-p)$.

Distribución binomial

140. Claramente $f(x) \geq 0$. Además por el teorema del binomio $\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p+1-p)^n = 1$.

141. La esperanza es $E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$. La suma puede empezar desde uno. Factorizando np y escribiendo $n-x = (n-1)-(x-1)$ se encuentra la expresión

$$np \sum_{x=1}^{n} \frac{(n-1)!}{((n-1)-(x-1))!(x-1)!} p^{x-1} (1-p)^{(n-1)-(x-1)},$$

lo cual es $np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{(n-1)-(x-1)}$. Haciendo el cambio de variable $y = x-1$ en la suma, ésta se convierte en la suma completa de la distribución $\mathrm{bin}(n-1, p)$ y vale uno. El resultado es entonces np. Para la varianza se usa la misma técnica. Se calcula primero el segundo momento escribiendo $x^2 = x(x-1) + x$, es decir,

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x (1-p)^{n-x} + \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}. \end{aligned}$$

La segunda suma acaba de ser calculada y vale np. La primera suma puede empezar desde 2, y procediendo como en el caso de la esperanza puede escribirse como

$$n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{((n-2)-(x-2))!(x-2)!} p^{x-2} (1-p)^{(n-2)-(x-2)},$$

lo cual es $n(n-1)p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} (1-p)^{(n-2)-(x-2)}$. Haciendo el cambio de variable $y = x-2$ en la suma, ésta se convierte en la suma completa de la distribución $\text{bin}(n-2, p)$ y vale uno. El segundo momento es entonces $n(n-1)p^2 + np$. Por lo tanto $\text{Var}(X) = E(X^2) - E^2(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p)$.

142. $n = 8$ y $p = 1/2$.

143. $P(Y = x) = P(n - X = x) = P(X = n - x)$. Al substituir la función de probabilidad de X resulta que Y tiene distribución binomial con el mismo número de ensayos n, pero la probabilidad de éxito es ahora $1-p$. Si X cuenta el número de éxitos en n ensayos Bernoulli, entonces Y cuenta el número de fracasos.

144. Desarrolle el lado derecho.

145. $\binom{6}{3}(1/2)^6$.

146. $\binom{2n}{n}(1/2)^{2n}$.

147. Recuerde que $\text{Var}(X) = npq$ y $E(X) = np$. Entonces efectivamente $npq \leq np$ pues $q \leq 1$.

148. Los tres eventos tienen la misma probabilidad de ocurrir, cada uno de ellos tiene probabilidad $1/2^6$.

Distribución geométrica

149. $f(x) \geq 0$. Además $\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} p(1-p)^x = p(1/p) = 1$.

150. Usaremos resultados de sumas geométricas. $E(X) = \sum_{x=0}^{\infty} xp(1-p)^x = \sum_{x=1}^{\infty}\sum_{y=1}^{x} p(1-p)^x$. Intercambiando el orden de las sumas se encuentra la expresión $\sum_{y=1}^{\infty}\sum_{x=y}^{\infty} p(1-p)^x = \sum_{y=1}^{\infty}(1-p)^y = (1-p)/p$. Para la varianza encontraremos primero el segundo momento, y para ello usaremos la identidad $x^2 = x(x+1) - x$. $E(X^2) = \sum_{x=1}^{\infty} x^2 p(1-p)^x = \sum_{x=0}^{\infty} x(x+1)p(1-p)^x - \sum_{x=0}^{\infty} xp(1-p)^x$. La segunda suma acaba de ser calculada y vale (1-p)/p. Para la primera suma escriba $x(x+1)$ como $2\sum_{y=1}^{x} y$. Intercambiando ahora el orden de las sumas se llega a $E(X^2) = 2\sum_{y=1}^{\infty} y \sum_{x=y}^{\infty} p(1-p)^x - (1-p)/p$ $= 2\sum_{y=1}^{\infty} y(1-p)^y - (1-p)/p = \frac{2}{p}\sum_{y=1}^{\infty} yp(1-p)^y - (1-p)/p = 2(1-p)/p^2 - (1-p)/p$. Por lo tanto la varianza es $\text{Var}(X) = E(X^2) - E^2(X)$ $= 2(1-p)/p^2 - (1-p)/p - (1-p)^2/p^2 = (1-p)/p^2$.

151. $P(X \geq a+b \,|\, X \geq a) = P(X \geq a+b)/P(X \geq a) = p\sum_{x=a+b}^{\infty}(1-p)^x/(p\sum_{x=a}^{\infty}(1-p)^x) = (1-p)^{a+b}/(1-p)^a = (1-p)^b = P(X \geq b)$.

152. 8/3=2.66 extracciones.

Distribución Poisson

153. La función $f(x)$ es no negativa y es tal que $\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!}$ $= e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$. Para la esperanza tenemos que $E(X) = \sum_{x=0}^{\infty} x\, e^{-\lambda}\frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} e^{-\lambda}\frac{\lambda^x}{(x-1)!} = \lambda\sum_{x=1}^{\infty} e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!} = \lambda$. Usando la igualdad $x^2 = x(x-1) + x$, puede encontrarse que el segundo momento es $E(X^2) = \lambda^2 + \lambda$. Por lo tanto Var($X$)$= E(X^2) - E^2(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

154. Desarrolle el lado derecho.

155. Sume las series $e^{\lambda} = 1 + \lambda + \lambda^2/2! + \lambda^3/3! + \cdots$, y $e^{-\lambda} = 1 - \lambda + \lambda^2/2! - \lambda^3/3! + \cdots$

156. a) 0.323 b) 0.593 c) Una o dos computadoras descompuestas tienen probabilidad máxima 0.2706.

Distribución binomial negativa

157. Este ejercicio no es fácil. Una forma de resolverlo es a través de la fórmula $\binom{-n}{k} = (-1)^k \binom{n+k-1}{k}$, y la expansión $(1+t)^a = \sum_{x=0}^{\infty} \binom{a}{x} t^x$ para a real y $|t| < 1$. Entonces $\sum_{x=0}^{\infty} \binom{r+x-1}{x} p^r (1-p)^x = p^r \sum_{x=0}^{\infty} (-1)^x \binom{-r}{x} (1-p)^x$ $= p^r \sum_{x=0}^{\infty} \binom{-r}{x} (p-1)^x = p^r (1+p-1)^{-r} = 1$.

158. Para la esperanza, factorice el término $r(1-p)/p$ en la expresión $E(X) = \sum_{x=1}^{\infty} x \binom{r+x-1}{x} p^r (1-p)^x$. Observe que la suma puede empezar desde 1. La suma resultante es uno pues se trata de las probabilidades de la distribución bin neg(r,p). Para la varianza use la fórmula $x^2 = x(x-1) + x$ para calcular primero $E(X^2)$, separando las sumas. El procedimiento es análogo al caso de la esperanza, el resultado es $E(X^2) = r(r+1)(1-p)^2/p^2 + r(1-p)/p$. Al substituir estos resultados en la fórmula $\text{Var}(X) = E(X^2) - E^2(X)$ se obtiene $\text{Var}(X) = r(1-p)/p^2$.

159. Factorice el término $(1-p)(r+x-1)/x$ en la expresión de $P(X=x)$.

Distribución hipergeométrica

160. El resultado se sigue de la fórmula $\sum_{k=0}^{r} \binom{n}{k} \binom{m}{r-k} = \binom{n+m}{r}$.

Distribución uniforme continua

161. Claramente $f(x) \geq 0$, y la integral de $f(x)$ es el área del rectángulo de base $b-a$ y altura $1/(b-a)$ que vale uno.

162. a) $E(X) = \int_a^b x\, 1/(b-a)\, dx = (b^2 - a^2)/(2(b-a)) = (a+b)/2$. b) $E(X^2) = \int_a^b x^2\, 1/(b-a)\, dx = (b^3 - a^3)/(3(b-a)) = (a^2 + ab + b^2)/3$. Por lo tanto $\text{Var}(x)$ $= E(X^2) - E^2(X) = (a^2 + ab + b^2)/3 - (a+b)^2/4 = (b-a)^2/12$.

163. $E(X^n) = \int_0^1 x^n\, 1\, dx = 1/(n+1)$.

164. $E(X^n) = \int_a^b x^n\, 1/(b-a)\, dx = \frac{1}{(n+1)(b-a)} x^{n+1} \Big|_a^b = (b^{n+1} - a^{n+1})/((n+1)(b-a))$.

165. $E(X^n) = \int_{-1}^{1} x^n \, 1/2 \, dx = \frac{1}{2(n+1)} \, x^{n+1}\Big|_{-1}^{1}$. Si n es impar, entonces $n+1$ es par y la integral se anula. Si n es par, entonces $n+1$ es impar y la integral es $1/(n+1)$.

166. X toma valores en $(0,1)$ y Y toma valores en $(-5,5)$. Para cualquier $y \in (-5,5)$, $F_Y(y) = P(Y \leq y) = P(10X - 5 \leq y) = P(X \leq (y+5)/10) = F_X((y+5)/10)$. Derivando $f_Y(y) = f_X((y+5)/10)\, 1/10$. Pero $f_X((y+5)/10)$ vale uno sólo cuando $0 < (y+5)/10 < 1$, es decir, si $-5 < y < 5$. Por lo tanto, Y tiene distribución unif$(-5,5)$.

167. La mediana en este caso es única y vale $m = (a+b)/2$ pues cumple la condición $P(X \leq m) = P(X \geq m) = 1/2$. Vea un ejercicio anterior para la demostración de que la media es también $(a+b)/2$.

Distribución exponencial

168. $f(x) \geq 0$ y $\int_0^\infty \lambda\, e^{-\lambda x}\, dx = -e^{-\lambda x}|_0^\infty = 1$.

169. $E(X) = \int_0^\infty x\, \lambda\, e^{-\lambda x}\, dx$. Usando integración por partes con $u = x$ y $dv = \lambda\, e^{-\lambda x}\, dx$, se obtiene $E(X) = \int_0^\infty e^{-\lambda x}\, dx = (1/\lambda) \int_0^\infty \lambda\, e^{-\lambda x}\, dx = 1/\lambda$. Para la varianza calcule primero el segundo momento, usando nuevamente integración por partes, $E(X^2) = \int_0^\infty x^2\, \lambda\, e^{-\lambda x}\, dx = 2\int_0^\infty x\, e^{-\lambda x}\, dx = (2/\lambda) \int_0^\infty x\, \lambda\, e^{-\lambda x}\, dx = 2/\lambda^2$. Por lo tanto $\text{Var}(X) = E(X^2) - E^2(X) = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$.

170. $P(X \geq x+y \,|\, X \geq x) = P(X \geq x+y)/P(X \geq x) = e^{-\lambda(x+y)}/e^{-\lambda x} = e^{-\lambda y} = P(X \geq y)$.

171. $F(x)\, F(y) = (1-e^{-\lambda x})(1-e^{-\lambda y}) = 1 - e^{-\lambda x} - e^{-\lambda y} + e^{-\lambda(x+y)} = (1-e^{-\lambda x}) + (1-e^{-\lambda y}) - (1-e^{-\lambda(x+y)}) = F(x) + F(y) - F(x+y)$.

172. Suponga que el tiempo se mide en minutos. $E(X) = 10 = 1/\lambda$. Por lo tanto $\lambda = 1/10$. Entonces $P(X > 60) = 1 - P(X \leq 60) = 1 - (1 - e^{-60/10}) = e^{-6} = 0.002478$. Por lo tanto de 1000 usuarios, aproximadamente 2.47 tienen conexión mayor a una hora. a) $P(X < 10) = F(10) = 1 - e^{-10/10} = 1 - e^{-1} = 0.6321$. b) $P(10 < X < 60) = F(60) - F(10) = (1-e^{-6}) - (1-e^{-1}) = 0.3654$.

Distribución gama

173. La función es no negativa, además haciendo el cambio de variable $t = \lambda x$ se obtiene $\int_0^\infty \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x}\, dx = \frac{1}{\Gamma(n)} \int_0^\infty t^{n-1} e^{-t}\, dt = \frac{\Gamma(n)}{\Gamma(n)} = 1$.

174. a) $E(X) = \int_0^\infty x \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x}\, dx = \frac{\Gamma(n+1)}{\lambda \Gamma(n)} \int_0^\infty \frac{(\lambda x)^n}{\Gamma(n+1)} \lambda e^{-\lambda x}\, dx$. La integral vale uno, y usando la propiedad $\Gamma(n+1) = n\,\Gamma(n)$ se obtiene $E(X) = n/\lambda$.
b) $E(X^2) = \int_0^\infty x^2 \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x}\, dx = \frac{\Gamma(n+2)}{\lambda^2 \Gamma(n)} \int_0^\infty \frac{(\lambda x)^{n+1}}{\Gamma(n+2)} \lambda e^{-\lambda x}\, dx = (n+1)n/\lambda^2$. Por lo tanto $\text{Var}(X) = E(X^2) - E^2(X) = (n+1)n/\lambda^2 - n^2/\lambda^2 = n/\lambda^2$.
c) Análogo al inciso anterior.

175. a) $\Gamma(n+1) = \int_0^\infty t^n e^{-t}\, dt$. Use integración por partes con $u = t^n$ y $dv = e^{-t}\, dt$. b) Por el inciso anterior, $\Gamma(2) = 1\,\Gamma(1)$, por lo tanto es suficiente demostrar que $\Gamma(1) = 1$, pero $\Gamma(1) = \int_0^\infty e^{-t}\, dt = 1$. c) Esto es consecuencia de los dos incisos anteriores. d) Haciendo el cambio de variable $x^2/2 = t$, $\Gamma(1/2) = \int_0^\infty t^{-1/2} e^{-t}\, dt = \sqrt{2} \int_0^\infty e^{-x^2/2}\, dx = \sqrt{2}\sqrt{2\pi}\frac{1}{2} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \sqrt{\pi}$. El último integrando es la densidad normal estándar.

Distribución beta

176. Claramente $f(x) \geq 0$, y $\int_0^1 \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}\, dx = \frac{1}{B(a,b)} \int_0^1 x^{a-1} (1-x)^{b-1}\, dx = \frac{B(a,b)}{B(a,b)} = 1$.

177. a) $E(X) = \int_0^1 \frac{x}{B(a,b)} x^{a-1} (1-x)^{b-1}\, dx = \frac{B(a+1,b)}{B(a,b)} \int_0^1 \frac{1}{B(a+1,b)} x^a (1-x)^{b-1}\, dx = \frac{B(a+1,b)}{B(a,b)}$. Ahora se usa la identidad $B(a+1,b) = \frac{a}{a+b} B(a,b)$ para obtener $E(X) = a/(a+b)$.
b) $E(X^2) = \int_0^1 \frac{x^2}{B(a,b)} x^{a-1} (1-x)^{b-1}\, dx = \frac{B(a+2,b)}{B(a,b)} \int_0^1 \frac{1}{B(a+2,b)} x^{a+1} (1-x)^{b-1}\, dx = \frac{B(a+2,b)}{B(a,b)}$. En donde $B(a+2,b) = \frac{a+1}{a+b+1} B(a+1,b) = \frac{a+1}{a+b+1} \frac{a}{a+b} B(a,b)$. Por lo tanto $E(X^2) = \frac{a+1}{a+b+1} \frac{a}{a+b}$, y entonces $\text{Var}(X) = \frac{a+1}{a+b+1} \frac{a}{a+b} - \frac{a^2}{(a+b)^2} = \frac{ab}{(a+b+1)(a+b)^2}$.

178. a) Efectúe el cambio de variable $y = 1 - x$.
b) $B(a,1) = \int_0^1 a x^{a-1}\, dx = 1/a$.
c) $B(1,b) = \int_0^1 a(1-x)^{b-1}\, dx = 1/b$.

d) $B(a+1,b)=\int_0^1 x^a(1-x)^{b-1}\,dx$. Use integración por partes con $u=x^a$ y $dv=(1-x)^{b-1}\,dx$ para llegar a $B(a+1,b)=\frac{a}{b}\int_0^1 x^{a-1}(1-x)^b\,dx$= $\frac{a}{b}B(a,b+1)$.
e) Puede usarse la identidad $B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)$. $B(a+1,b)=\Gamma(a+1)\Gamma(b)/\Gamma(a+b+1)=\frac{a}{a+b}\Gamma(a)\Gamma(b)/\Gamma(a+b)$= $\frac{a}{a+b}B(a,b)$.
f) Por el inciso anterior, $B(a,b+1)$= $B(b+1,a)$= $\frac{b}{a+b}B(b,a)$= $\frac{b}{a+b}B(a,b)$.
g) Efectúe el cambio de variable $x=\cos^2\theta$.

179. Simplemente sustituya esos valores para los parámetros a y b en la distribución beta(a,b). Recuerde que $B(1,1)=1$.

180. a) $f(x)=1/(\pi\sqrt{x(1-x)}$, para $0<x<1$.
b) Efectúe el cambio de variable $x=\cos^2\theta$.
c) $E(X)=B(1+1/2,1/2)/B(1/2,1/2)=(1/2)/(1/2+1/2)=1/2$. Para la varianza use $E(X^2)=B(2+1/2,1/2)/B(1/2,1/2)$.

181. Compruebe que $B(a,1)=1/a$. Por lo tanto $F(x)$= $(\int_0^x u^{a-1}du)/B(a,1)$= x^a, para $x\in(0,1)$.

182. Compruebe que $B(1,b)=1/b$. Por lo tanto $F(x)$= $(\int_0^x(1-u)^{b-1}du)/B(1,b)$= $1-(1-x)^b$, para $x\in(0,1)$.

183. $E(X^n)$= $(\int_0^1 x^n\,x^{a-1}(1-x)^{b-1}dx)/B(a,b)=B(a+n,b)/B(a,b)$.

Distribución normal

184. Este ejercicio no es fácil pero puede resolverse por varios métodos. El siguiente método hace uso de coordenadas polares. Sea $I=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}\,dx$ $=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx$. Entonces I^2= $(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx)\,(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy)$ $=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-(x^2+y^2)/2}\,dx\,dy=\int_0^{2\pi}\int_0^{\infty}\frac{1}{2\pi}e^{-r^2/2}r\,dr\,d\theta$, en donde la última integral se obtiene después del cambio de variable a coordenadas polares (x,y)= $(r\cos\theta,r\operatorname{sen}\theta)$. Entonces I^2= $\int_0^{\infty}e^{-r^2/2}r\,dr$= 1.

185. Considere la esperanza $E(X)=\int_{-\infty}^{\infty}x\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}\,dx$. Haciendo el cambio de variable $u=(x-\mu)/\sigma$ se encuentra que $E(X)=\mu+\sigma E(Z)$, en donde $Z\sim\mathrm{N}(0,1)$. Por lo tanto es suficiente demostrar que $E(Z)=0$, pero ello es fácil pues el integrando de $E(Z)=\int_{-\infty}^{\infty}u\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du$ es una

derivada excepto por el signo. Por lo tanto $E(X) = \mu$. Para la varianza considere el segundo momento $E(X^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}\, dx$. Efectúe nuevamente el cambio de variable $y = (x-\mu)/\sigma$ y encuentre que $E(X^2)$= $\mu^2+2\mu E(Z)+\sigma^2 E(Z^2)$. Resta encontrar $E(Z^2) = \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$. Esta integral puede realizarse por partes con $u = y$ y $dv = y\frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$. El resultado es 1. Por lo tanto $E(X^2) = \mu^2+\sigma^2$, y entonces Var(X)= $(\mu^2+\sigma^2)-\mu^2$= σ^2.

186. Suponga $X \sim \mathrm{N}(\mu, \sigma^2)$. Sea $Z = (X-\mu)/\sigma$. Entonces $F_Z(u)$= $P(Z \leq u)$= $P(\frac{X-\mu}{\sigma} \leq u)$= $P(X \leq \mu+u\sigma)$= $F_X(\mu+u\sigma)$. Derivando respecto a u, $f_Z(u)$= $f_X(\mu+u\sigma)\sigma$. Substituya la expresión para $f_X(x)$ y encuentre la densidad normal estándar. El procedimiento es reversible. Suponga $Z \sim \mathrm{N}(0,1)$. Sea $X = \mu + \sigma Z$. Entonces $F_X(x)$= $P(X \leq x) = P(\mu+\sigma Z \leq x)$= $P(Z \leq (x-\mu)/\sigma)$= $F_Z((x-\mu)/\sigma)$. Derivando respecto a x, $f_X(x)$= $f_Z((x-\mu)/\sigma)/\sigma$. Substituya la expresión de $f_Z(x)$ y encuentre la densidad $\mathrm{N}(\mu, \sigma^2)$.

187. a) $P(X \geq 10)$= 1/2. b) $P(X < 0)$= $P(Z < -2)$= 0.0228. c) $P(0 < X \leq 10)$= $P(-2 < Z \leq 0)$= 0.4772. d) $P(X \geq 20)$= $P(Z \geq 2)$= 0.0228. e) $P(-20 < X \leq 10)$= $P(-6 < Z \leq 0)$= 1/2.

188. a) $P(X \leq 10)$= $P(Z \leq 1)$= 0.8413. b) $P(X > 0)$= 1/2. c) $P(0 < X \leq 40)$= $P(0 < Z \leq 4)$= 1/2. d) $P(X \geq 30)$= $P(Z \geq 3)$= 0.0013. e) $P(-10 < X \leq 10)$= $P(-1 < Z \leq 1)$= 0.6826.

189. a) $x = 1.115$. b) $x = -1.375$.

190. Suponga $a > 0$. Entonces $F_Y(y)$= $P(Y \leq y) = P(aX+b \leq y)$= $P(X \leq (y-b)/a)$= $F_X((y-b)/a)$. Derivando, $f_Y(y)$= $f_X((y-b)/a)/a$. Substituya ahora en esta expresión la función de densidad $\mathrm{N}(\mu, \sigma^2)$ y encuentre la función de densidad normal ahora con media $a\mu+b$ y varianza $a^2\sigma^2$. Haga un análisis semejante para el caso $a < 0$, tenga cuidado al dividir por tal cantidad. Los nuevos parámetros son idénticos al caso anterior.

191. Sea $Y = -X$. Entonces $F_Y(y)$= $P(Y \leq y)$= $P(-X \leq y) = P(X \geq -y)$= $1-P(X < -y)$= $1-F_X(-y)$. Derivando, $f_Y(y)$= $f_X(-y)$. Substituya ahora la función de densidad de X evaluada en $-y$ y compruebe que se trata de la densidad normal con media $-\mu$ y varianza σ^2.

192. Para $x > 0$, $F_{X^2}(x)$= $P(X^2 \leq x)$= $P(-\sqrt{x} \leq X \leq \sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x})$. Derivando, $f_{X_2}(x)$= $f_X(\sqrt{x})\frac{1}{2\sqrt{x}}+f_X(-\sqrt{x})\frac{1}{2\sqrt{x}} = f_X(\sqrt{x})\frac{1}{\sqrt{x}}$.

Ahora substituya la expresión para $f_X(x)$ y encuentre la función de densidad de la distribución $\chi^2(1)$.

193. Para $y > 0$, $F_Y(y)= P(Y \leq y)= P(|X| \leq y)= P(-y \leq X \leq y)= F_X(y) - F_X(-y)$. Derivando, $f_Y(y)= f_X(y)+f_X(-y)= 2f_X(y)$. Ahora substituya la función de densidad de X.

194. Sea X el llenado de un vaso cualquiera. Suponga $X \sim \mathrm{N}(320, 100)$. Entonces a) $P(X > 310)= P(Z > 1)= 0.1587$. b) $P(290 < X < 305)= P(-1 < Z < 1/2)= 0.5328$. c) $P(X > 320)= P(Z > 2)= 0.0228$. d) $P(X < 270)= P(Z < -3)= 0.0013$. De mil clientes, 1.3 reclamarán por vasos servidos con 270 ml. o menos.

Distribución ji cuadrada

195. La función es no negativa. Efectúe el cambio de variable $t = x/2$ dentro de la integral. La integral resultante es la función gama evaluada en $n/2$.

196. simplemente substituya el valor $n = 2$ en la densidad $\chi^2(n)$. Recuerde que $\Gamma(1) = 1$.

197. En estos cálculos es conveniente hacer el cambio de variable $t = x/2$ en cada una de las integrales. Para la esperanza, el exponente de t es $n/2$, que puede ser escrito como $(n/2) + 1 - 1$. Reconstruya la integral como la función gama evaluada en $(n/2) + 1$. Después simplifique usando la propiedad $\Gamma(n + 1) = n\,\Gamma(n)$. Para la varianza, calcule primero el segundo momento, el exponente de t es $(n/2)+1$, que puede ser escrito como $(n/2)+2-1$. Ahora reconstruya la integral resultante como la función gama evaluada en $(n/2) + 2$. Después simplifique. El cálculo del m-ésimo momento sigue las mismas líneas.

Distribución t

198. La función es no negativa. Para comprobar que esta función integra uno efectúe el cambio de variable $1 - u = (1 + x^2/n)^{-1}$ y reconstruya la integral de $B(1/2, n/2)$.

199. La integral $E(X)$ es absolutamente convergente y el integrando es una función impar, por lo tanto la integral es cero. Dado este resultado, la varianza es el

segundo momento. Efectúe el cambio de variable $1-u=(1+x^2/n)^{-1}$ en la integral $E(X^2)$, y reconstruya la integral de $B(3/2,(n-2)/2)$, después simplifique.

Vectores aleatorios

200. —

201. —

202. —

203. —

Variables y tipos de datos

204. —

Estadística descriptiva

205. —

206. —

207. —

208. $\sum_{i=1}^n (x_i-\bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = \sum_{i=1}^n x_i - \sum_{i=1}^n x_i = 0.$

209. $\sum_{i=1}^n (x_i-\bar{x})^2 = \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + n\bar{x}^2 = \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2.$

210. Idéntico al ejercicio anterior.

211. Derive la función $\bar{x}(c)$ respecto de la variable c, iguale a cero y encuentre que $c=\bar{x}$. Compruebe que la segunda derivada es negativa.

212. —

213. —

Muestras aleatorias y estadísticas

214. —

Estimación puntual

215. $E(\hat{\theta}) = E(\alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2) = \alpha E(\hat{\theta}_1) + (1-\alpha)E(\hat{\theta}_2) = \alpha\theta + (1-\alpha)\theta = \theta.$

216. $E(\hat{\sigma}^2) = \dfrac{1}{n}\sum_{i=1}^{n} E(X_i - \mu)^2 = \dfrac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \dfrac{1}{n}\sum_{i=1}^{n} \sigma^2 = \sigma^2.$

217. —

218. $E(\hat{\sigma}^2) = \dfrac{1}{2(n-1)}\sum_{i=1}^{n-1} E(X_{i+1} - X_i)^2 = \dfrac{1}{2(n-1)}\sum_{i=1}^{n-1} (E(X_{i+1}^2) - 2E(X_{i+1})E(X_i) +$
$E(X_i^2)) = \dfrac{1}{2(n-1)}\sum_{i=1}^{n-1}(\sigma^2 + \mu^2 - 2\mu^2 + \sigma^2 + \mu^2) = \dfrac{1}{2(n-1)}\sum_{i=1}^{n-1} 2\sigma^2 = \sigma^2.$

219. —

Método de momentos

220. —

221. —

222. —

Método de máxima verosimilitud

223. —

224. $\hat{\lambda} = \bar{X}.$

Estimación por intervalos

225. —

226. —

Pruebas de hipótesis

227. Se acepta H_0.

228. Se rechaza H_0.

229. Cuando $n \to \infty$, el cociente $\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}$ se va a infinito o menos infinito dependiendo del signo de la diferencia $\mu_0 - \mu_1$.

Apéndice C

Formulario

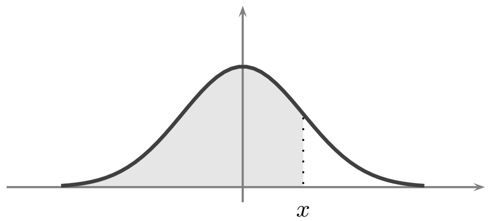
Notación

$\mathbb{N}$	Conjunto de números naturales $1, 2, 3, \ldots$
$\mathbb{Z}$	Conjunto de números enteros $0, \pm 1, \pm 2, \pm 3, \ldots$
$\mathbb{Q}$	Conjunto de números racionales a/b en donde $a, b \in \mathbb{Z}$ con $b \neq 0$.
$\mathbb{R}$	Conjunto de números reales.

El alfabeto griego

A α	alpha	I ι	iota	P ρ, ϱ	rho
B β	beta	K κ	kappa	Σ σ, ς	sigma
Γ γ	gamma	Λ λ	lambda	T τ	tau
Δ δ	delta	M μ	mu	Υ υ	upsilon
E ϵ, ε	epsilon	N ν	nu	Φ ϕ, φ	phi
Z ζ	zeta	Ξ ξ	xi	X χ	chi
H η	eta	O o	omikron	Ψ ψ	psi
Θ θ, ϑ	theta	Π π	pi	Ω ω	omega

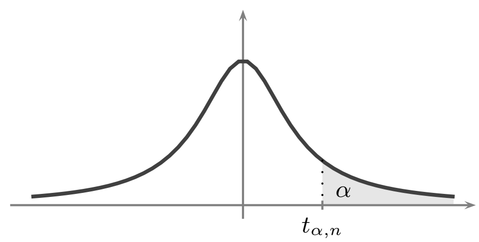
Tabla de la distribución normal estándar



$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8399
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Tabla de la distribución t(n)



$P(T \geq t_{\alpha,n}) = \alpha$

n	α =0.005	α =0.01	α =0.025	α =0.05	α =0.1
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.474	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.291	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
∞	2.576	2.326	1.960	1.645	1.282

Bibliografía

[1] Blake I. F. *An introduction to applied probability.* John Wiley & Sons, 1979.

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

[3] Devore J. *Probabilidad y estadística para ingeniería y ciencias.* Thomson, 2001.

[4] Feller W. *Introducción a la teoría de probabilidades y sus aplicaciones.* Limusa, 1973.

[5] Garza T. *Elementos de cálculo de probabilidades.* UNAM, 1983.

[6] Hernández-Del-Valle A. y Hernández-Lerma O. *Elementos de probabilidad y estadística.* Serie Textos No. 21, Nivel Elemental. Sociedad Matemática Mexicana, 2003.

[7] Hoel P. G., Port S. C., Stone C. J. *Introduction to probability theory.* Houghton Mifflin, 1971.

[8] Hoel P. G., Port S. C., Stone C. J. *Introduction to statistical theory.* Houghton Mifflin, 1971.

[9] Hogg R. V., Tanis E. A. *Probability and statistical inference.* 1993.

[10] Kolmogorov A. N. *Foundations of the theory of probability.* Chelsea Publishing Company, 1950.

[11] Mood A. M., Graybill F. A., Boes D. C. *Introduction to the theory of statistics.* McGraw Hill, 1983.

[12] Miller I., Miller M. *John E. Freund's mathematical statistics.* Prentice Hall, 1999.

[13] Ross S. *A first course in probability.* Macmillan Publishing Company, 1994.

[14] Ross S. M. *Introduction to probability and statistics for engineers and scientists.* 2da Edición. Academic Press, 2000.

[15] Wisniewski P. M., Bali G. *Ejercicios y problemas de la teoría de las probabilidades.* Trillas, 1998.

[16] Velasco G., Wisniewski P. *Probabilidad y estadística para ingeniería y ciencias.* Thomson, 2001.

Índice