Versión preliminar

Una introducción a la
PROBABILIDAD Y ESTADÍSTICA

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

El presente texto corresponde a la versión electrónica de agosto de 2006.
Este material se encuentra en permanente actualización y corrección.
La última versión disponible puede obtenerse en

`http://www.matematicas.unam.mx/lars`

Se sugiere imprimir por ambos lados,
cortar todas las hojas por la linea punteada y después encuadernar.
La caja mide 16cm por 21.5cm si se imprime sin reducción.

Prefacio

El presente texto constituye el material completo del curso semestral de *Probabilidad y Estadística*, impartido por el autor en la Facultad de Ciencias de la UNAM. Contiene el temario básico para un curso elemental e introductorio de probabilidad y estadística, así como una colección de ejercicios.

El texto está dirigido 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*.

Dado el caracter preliminar de este trabajo, el autor agradece cualquier comentario, sugerencia o corrección enviada al correo electrónico que aparece abajo.

Luis Rincón
Agosto 2006
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 sencillos, algunas personas los utilizan para tomar decisiones en sus vidas. 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 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). 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*. 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.

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. $\circ$

Si al lanzar un dado una vez obtenemos el número “4”, decimos entonces que se observó la ocurrencia del evento $A = \{2, 4, 6\}$, y si se obtiene por ejemplo el resultado “1” decimos que no se observó la ocurrencia del evento A.

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.

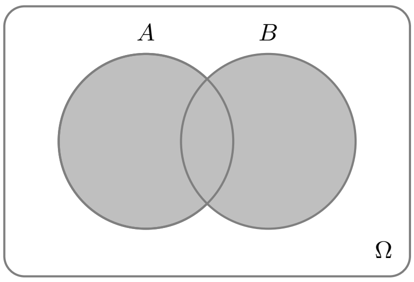
Conjuntos

Supondremos entonces que el espacio muestral Ω de un experimento aleatorio es nuestro 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 $\#A$.

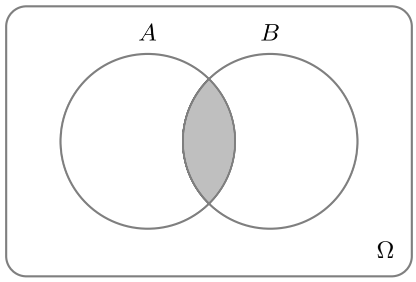
Sean A y B dos subconjuntos cualesquiera de Ω. Recordamos a continuación las operaciones básicas de unión, intersección, diferencia y complemento:

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

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". Mostramos a continuación en diagramas de Venn estas operaciones.

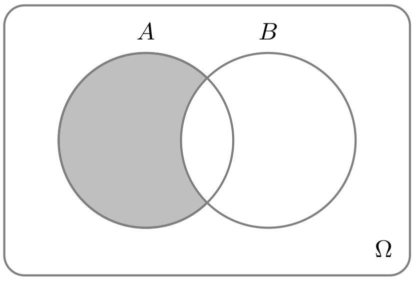


$A \cup B$

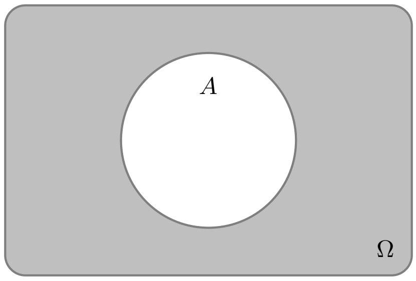


$A \cap B$

El complemento de un conjunto A se denota por A^c y se define como la colección de aquellos elementos de Ω que no pertenecen al conjunto A. Mediante un diagrama de Venn ilustramos gráficamente las operaciones de diferencia y complemento.



$A - B$



A^c

Es fácil verificar que el conjunto vacío $\emptyset$ y el conjunto total Ω satisfacen las siguientes propiedades elementales:

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

Las operaciones unión e intersección son *asociativas*, esto es, satisfacen las siguientes igualdades:

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

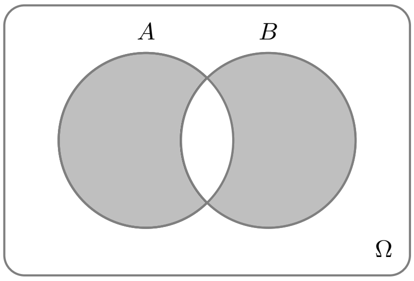
y también son *distributivas*, es decir,

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

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 siguiente figura 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 $(A - B) \cup (B - A)$.



$A \triangle B$

Recordemos además las *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 igualdades puede extenderse a colecciones finitas e incluso arbitrarias de 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, 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 $B = \{5, 6\}$ son ajenos pues no hay ningún elemento común entre ellos. Este concepto puede extenderse al caso de varios conjuntos de la siguiente forma: Decimos que n conjuntos $A_1, A_2, \ldots, A_n$ 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.

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,

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

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 el número de elementos en Ω. De este hecho proviene la notación usada para el conjunto potencia: 2^Ω. Para el ejemplo anterior se comprueba que efectivamente $\#(2^\Omega) = 2^{\#\Omega} = 2^3 = 8$.

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 $n \cdot m$. Más generalmente,

$$\#(A_1 \times A_2 \times \cdots \times A_n) = \#A_1 \cdot \#A_2 \cdots \#A_n.$$

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 experimentos aleatorios. En la siguiente sección estudiaremos algunas formas de definir matemáticamente la probabilidad de un evento cualquiera.

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 que 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 $A = \{2, 4, 6\}$, entonces

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

Probabilidad frecuentista

Supongamos que realizamos n veces 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

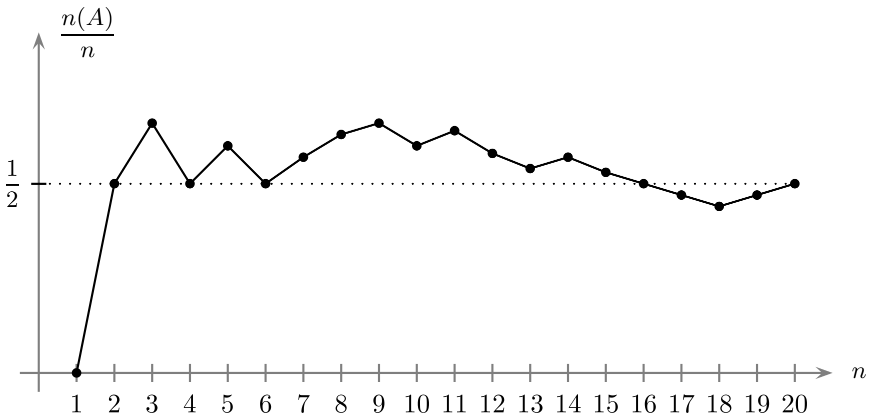
$$P(A) = \lim_{n\to\infty} \frac{n(A)}{n}.$$

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 obtuvimos 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 siguiente gráfica 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.



Regularidad estadística del cociente $\dfrac{n(A)}{n}$.

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 serio esta forma de definir 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 nuestro equipo favorito de futbol 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.

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 son tres postulados o axiomas[1] 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 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. Como consecuencia de estos postulados es posible demostrar que la probabilidad cumple, entre otras, con las siguientes propiedades.

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. $P(\emptyset) = 0$.

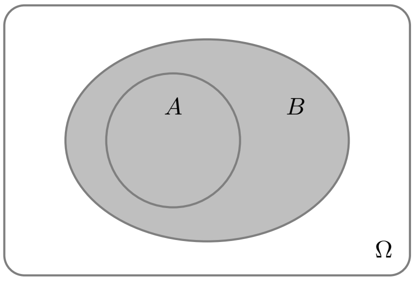
Demostración. Como $\emptyset = \Omega^c$, usando la propiedad anterior, tenemos que

$$P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 0.$$

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

□

Las siguientes dos proposiciones suponen la situación $A \subseteq B$ que se muestra gráficamente a continuación.



$A \subseteq B$

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 $P(B) = P(A) + P(B - A)$. □

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

Demostración. Como $A \subseteq \Omega$ entonces $P(A) \leq P(\Omega) = 1$. 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,

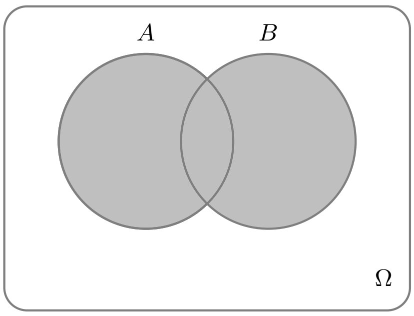
$$P(A \cup B) = P(A - A \cap B) + P(A \cap B) + P(B - A \cap B).$$

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

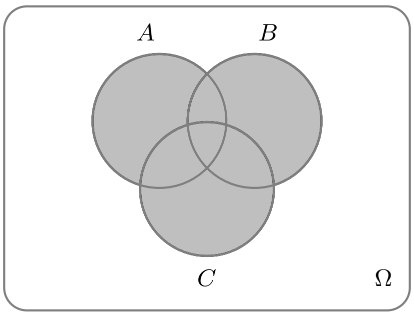
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

□

En la primera de las siguientes figuras 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 tomadas en cuenta una sola vez y el resultado es la probabilidad del evento $A \cup B$.



$A \cup B$



$A \cup B \cup C$

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 arriba. 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,

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

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

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

□

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

Es muy frecuente que en un experimento aleatorio el espacio muestral Ω sea un conjunto finito y cada elemento de este conjunto tenga la misma probabilidad de ocurrir, es decir, que el espacio Ω sea finito y equiprobable. En estos casos hemos definido la probabilidad clásica de un evento A como sigue:

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

Para poder aplicar esta definición necesitamos saber contar cuántos elementos tiene un conjunto A. 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? Estoy seguro que nadie en su sano juicio intentaría escribir uno a uno todos estos números telefónicos. En esta sección 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.

Principio de multiplicación

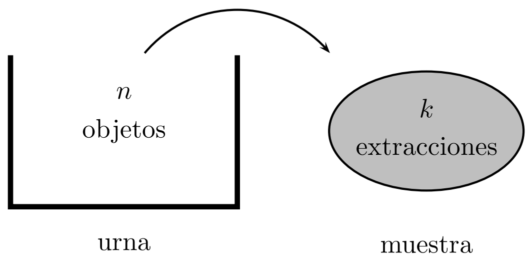
Si un procedimiento puede efectuarse de n formas distintas y un segundo procedimiento puede realizarse de m formas diferentes, entonces el total de formas en que puede efectuarse el primer procedimiento seguido del segundo es el producto $n \cdot m$. Para ilustrar el principio de multiplicación considere el siguiente ejemplo.

Ejemplo. Un 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 el principio de multiplicación se puede enunciar en símbolos de la forma siguiente:

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

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 siguiente figura:



Ordenaciones con repetición: Muestras con orden y con reemplazo

Suponga 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 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*. Veamos un ejemplo.

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. $\circ$

Ordenaciones sin repetición:
Muestras con orden y sin reemplazo

A veces no queremos ordenar todos los n objetos de un conjunto sino únicamente k de ellos ($k \leq n$) y sin repetirlos. La respuesta al total de arreglos lineales que podemos 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 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*.

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 $10 \times 9 \times 8 = 720$ distintas asignaciones de los tres primeros lugares en la rifa. $\circ$

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 $0! = 1$.

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 multiplicativo la respuesta es entonces el producto de estos números. ◦

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 seguimos el razonamiento siguiente. 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 estoy seguro 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. ¿Cuántas equipos distintos de tres personas pueden formarse de un grupo de 5 personas? Observe que el orden de las tres personas escogidas no importa de modo que la respuesta es

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = 10.$$

$\circ$

El coeficiente binomial es también una forma de generar las entradas del así llamado *triángulo de Pascal*, que puede observarse en la siguiente figura:

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}$$

Primeros renglones del triángulo de Pascal.

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 67 en la página 102.

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 asi sucesivamente, hasta k_m objetos del tipo m, en donde $k_1 + k_2 + \cdots + k_m = n$. Entonces estos n objetos pueden todos ordenarse uno detrás de otro de tantas formas distintas como indica el así llamado *coeficiente multinomial*:

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

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:

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

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 $(a + b + c)^3$?

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 siguiente arreglo de n casillas junto con la siguiente interpretación.

$$\begin{array}{|c|c|c|c|c|c|c|}
\times\times & & \times & \times & \cdots & & \times \\
\hline
\multicolumn{1}{c}{1} & \multicolumn{1}{c}{2} & \multicolumn{1}{c}{3} & \multicolumn{1}{c}{4} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{n-1} & \multicolumn{1}{c}{n}
\end{array}$$

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 parades 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.

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

Nota importante. 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.

1.4. Probabilidad condicional e independencia

En esta sección se estudian los conceptos importantes de probabilidad condicional e independencia. Estos conceptos surgieron de manera natural en el proceso de encontrar solución a algunos problemas provenientes de situaciones reales. Se demuestran 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 $P(B) > 0$. 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 entonces "probabilidad condicional del evento A dado el evento B" o simplemente "probabilidad de A dado B". Para que la definición tenga sentido se necesita suponer que $P(B) > 0$. No se define $P(A|B)$ cuando $P(B) = 0$. El evento B representa información adicional acerca del experimento aleatorio. Mediante un ejemplo sencillo se ilustra a continuación el uso y significado de la probabilidad condicional.

Ejemplo. Considere el experimento de lanzar un dado equilibrado. Claramente el espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$ el cual es equiprobable. Sean los eventos $A = \{2\}$ y $B = \{2, 4, 6\}$ = "Cae par". Entonces $P(A) = 1/6$ mientras que $P(A|B) = 1/3$. Observe que conocer la información de la ocurrencia del evento B, ha afectado la probabilidad del evento A. ◦

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

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

La demostración de esta fórmula es sencilla pues simplemente se escribe la definición de cada probabilidad condicional en el lado derecho, se cancelan términos y lo que resulta es el lado izquierdo.

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. 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:

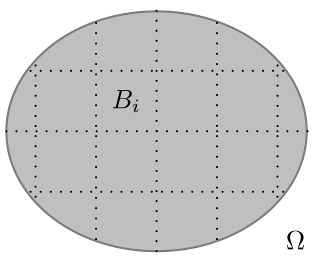
$$
\begin{aligned}
P(A_i \cap A_j) &= P(A_i)P(A_j), \quad i, j \text{ distintos.} \\
P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k), \quad i, j, k \text{ distintos.} \\
&\vdots \\
P(A_1 \cap \cdots \cap A_n) &= P(A_1) \cdots P(A_n).
\end{aligned}
$$

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?

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 $B_1, B_2, \ldots, B_n$ 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 siguiente mostramos gráficamente el concepto de partición de un conjunto.

$B_1, B_2, \ldots, B_n$
es una partición
del conjunto Ω.



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 $P(B_i) > 0$. Sea A cualquier evento. Entonces*

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

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 $A \cap B_i$, para $i = 1, 2, \ldots, n$, son ajenos. De modo que

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

□

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 sencilla:

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

Consideraremos a continuación algunos ejemplos de aplicación del teorema de probabilidad total.

Ejemplo. Supongamos que tenemos dos cajas: una con 3 bolas de color rojo y 7 de color negro, la otra con 6 rojas y 6 negras. Si se elije una caja al azar y después se saca una bola, ¿cuál es la probabilidad de que sea de color rojo? El experimento aleatorio consiste entonces en escoger una caja al azar y despues escoger una bola de la caja escogida. Es claro entonces que el espacio muestral puede escribirse como sigue

$$\Omega = \{(C_1, R), (C_1, N), (C_2, R), (C_2, N)\},$$

en donde C_1 y C_2 denotan los eventos en donde las cajas uno y dos fueron escogidas, respectivamente, y R y N denotan los eventos en donde una bola roja y negra fueron escogidas respectivamente. Nos piden calcular la probabilidad de R. Es fácil calcular la probabilidad de este evento cuando sabemos cuál caja fue escogida. Esto sugiere entonces condicionar sobre el resultado de escoger alguna de las dos cajas y aplicar el teorema de probabilidad total, es decir,

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

Observe que la partición de Ω consta de dos elementos: $\{(C_1, R), (C_1, N)\}$ y $\{(C_2, R), (C_2, N)\}$. ◦

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

$$\begin{aligned} M &= \text{"La persona escogida es mujer."} \\ H &= \text{"La persona escogida es hombre."} \\ D &= \text{"La persona escogida es daltónica."} \end{aligned}$$

Deseamos calcular $P(D)$. Por el teorema de probabilidad total,

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

◦

Teorema de Bayes

Otro resultado interesante que involucra probabilidades condicionales es el famoso teorema de Bayes. Este resultado 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 $B_1, B_2, \ldots, B_n$ una partición de Ω tal que $P(B_i) > 0$, y 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

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

□

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 simple:

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

Ejemplo. En una fábrica hay dos máquinas, que denotaremos por $\mathcal{A}$ y $\mathcal{B}$. La máquina $\mathcal{A}$ realiza el 60 % de la producción total y la máquina $\mathcal{B}$ el 40 %. De su producción, la máquina $\mathcal{A}$ produce 3 % de material defectuoso, la $\mathcal{B}$ el 5 %. Se ha encontrado un material defectuoso, ¿cuál es la probabilidad de que este material defectuoso provenga de la máquina $\mathcal{B}$? Sean los eventos

$$\begin{aligned} A &= \text{“La máquina } \mathcal{A} \text{ produjo el material escogido.”} \\ B &= \text{“La máquina } \mathcal{B} \text{ produjo el material escogido.”} \\ D &= \text{“El material escogido es defectuoso.”} \end{aligned}$$

Nos preguntan $P(B|D)$ y obervamos que la información que tenemos es $P(D|B)$. Por el teorema de Bayes tenemos entonces que

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

$\circ$

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,

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

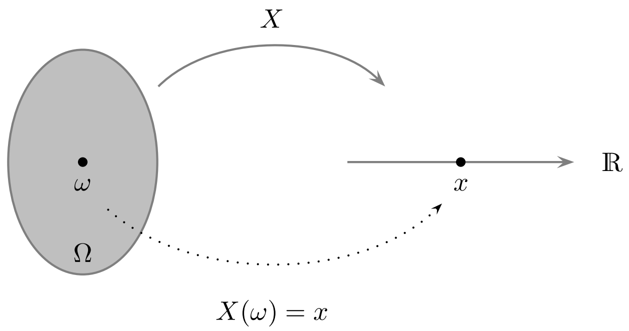
Esta última probabilidad es demasiado pequeña y por lo tanto la prueba no es muy confiable en tales casos. ∘

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 siguiente figura.



Una variable aleatoria es una función X del conjunto Ω en $\mathbb{R}$.

Debemos hacer aqui varias observaciones. Primeramente seguiremos la notación usual de usar la letra mayúscula X para denotar de manera general una variable aleatoria cualquiera. Es importante observar que X (mayúscula), denota una variable aleatoria, es decir, una función de Ω en $\mathbb{R}$, mientras que x (minúscula), denota un número real. Veamos algunos ejemplos sencillos.

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" y "Cruz" los dos lados de la moneda. Entonces claramente el espacio muestral es el conjunto $\Omega = \{\text{“Cara”}, \text{“Cruz”}\}$. Defina la variable aleatoria $X : \Omega \to \mathbb{R}$ como sigue

$$\begin{aligned} X(\text{“Cara”}) &= 0, \\ X(\text{“Cruz”}) &= 1. \end{aligned}$$

De este modo podemos suponer entonces que el experimento aleatorio tiene dos valores numéricos posibles: 0 y 1. Observe que los números 0 y 1 son en realidad arbitrarios y bien pueden ser escogidos otro par de números reales. ∘

Ejemplo. Considere nuevamente el experimento aleatorio sencillo de lanzar una moneda. Podemos definir otra variable aleatoria $Y : \Omega \to \mathbb{R}$ de la siguiente forma

$$Y(\text{“Cara”}) = Y(\text{“Cruz”}) = 2.$$

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 entonces 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 $\Omega = \{(x,y) : x^2+y^2 \leq 1\}$. Definiremos a continuación varias variables aleatorias, es decir funciones de Ω en $\mathbb{R}$, 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).

Observe que cada uno de estos ejemplos es una función de Ω en $\mathbb{R}$, y por lo tanto cada una de estas funciones es una variable aleatoria. ∘

Ahora, si consideramos el conjunto de valores que una variable aleatoria puede tomar, podemos clasificar las variables aleatorias en al menos dos tipos: discretas y 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.

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 subconjunto de Ω cuyos elementos son tales que bajo la aplicación de la función X toman un valor numérico contenido en el conjunto A.

Ejemplo. Consideramos nuevamente el ejemplo anterior de lanzar una moneda. Tenemos por ejemplo que $(X \in [1, \infty)) = \{\text{“Cruz”}\}$ pues el conjunto de elementos de Ω (solo hay dos elementos en Ω) tales que bajo la función X toman un valor numérico 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) $(X \in [1, 2)) = \{\text{“Cruz”}\}$, por lo tanto $P(X \in [1, 2)) = 1/2$.

b) $(X \in [0, 1)) = \{\text{“Cara”}\}$, por lo tanto $P(X \in [0, 1)) = 1/2$.

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

d) $(X = 1) = \{\text{“Cruz”}\}$, por lo tanto $P(X = 1) = 1/2$.

e) $(X \leq -1) = \emptyset$, por lo tanto $P(X \leq -1) = 0$.

f) $(X \geq 0) = \Omega$, por lo tanto $P(X \geq 0) = 1$.

$\circ$

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 $P(X \leq x) = 1 - P(X > x)$.

Nota importante. A través de una variable aleatoria podemos considerar ahora que los posibles resultados de un experimento aleatorio no son elementos ω en Ω sino números reales que la variable aleatoria puede tomar. Ya no consideraremos los eventos (subconjuntos) de Ω sino que ahora nuestros eventos serán subconjuntos de $\mathbb{R}$. De este modo la probabilidad del evento intervalo (a, b) es la probabilidad $P(X \in (a, b))$. A partir de ahora y en lo que resta del curso el término *variable aleatoria* constituirá un elemento bastante frecuente en nuestros 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 densidad* de la variable X denotada por $f(x) : \mathbb{R} \to [0, \infty)$ se define como sigue

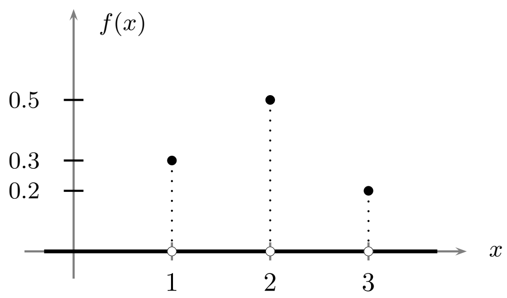
$$f(x) = \begin{cases} P(X = x) & \text{si } x = x_1, x_2, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

Recordemos que es importante poder distinguir entre X y x, pues conceptualmente son cosas muy distintas. Denotaremos generalmente a una función de densidad con la letra f minúscula de modo que el subíndice X nos ayuda a determinar que $f_X(x)$ es la función de densidad de la variable X. Esta notación será particularmente útil cuando consideremos varias variables aleatorias a la vez.

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 densidad de X es

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

Gráficamente,



Gráfica de una función de probabilidad.

Alternativamente podemos también expresar esta función mediante la siguiente tabla:

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

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 usted que las siguientes probabilidades son correctas.

$$\begin{aligned} P(X \geq 2) &= P(X = 2) + P(X = 3) = 0.7, \\ P(|X| = 1) &= P(X = 1) = 0.3, \\ P(X < 1) &= 0. \end{aligned}$$

∘

Función de densidad para una variable continua

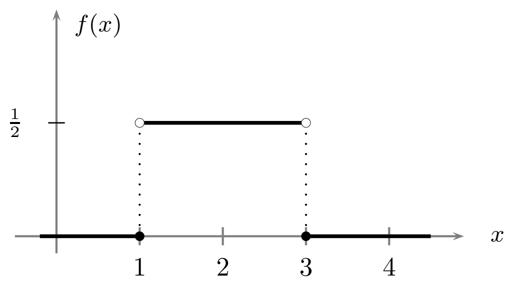
Sea X una variable aleatoria continua. Decimos que la función integrable y no negativa $f(x) : \mathbb{R} \to [0, \infty)$ es la función de densidad de X si para cualquier intervalo (a, b) de $\mathbb{R}$ se cumple la igualdad

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

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. Por 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 cuya gráfica aparece en la siguiente figura:



Ejemplo de una función de densidad.

No es difícil comprobar que toda función de densidad $f(x)$ de una variable aleatoria continua cumple las siguientes propiedades:

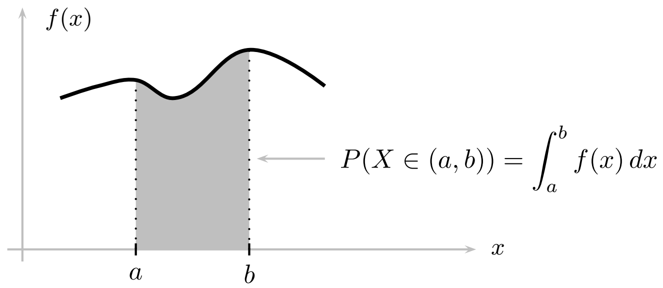
a) $f(x) \geq 0$, para toda $x \in \mathbb{R}$.

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

Estas dos propiedades se obtienen realmente de la definición de función de densidad para una variable continua. Efectivamente a $f(x)$ se le pide ser no negativa en la definición, además

$$\int_{-\infty}^{\infty} f(x)\,dx = P(X \in \mathbb{R}) = P(\Omega) = 1.$$

Toda función $f(x) : \mathbb{R} \to [0, \infty)$ que satisfaga las dos propiedades de la proposición anterior, sin necesidad de tener una variable aleatoria de por medio, se llamará función de densidad. Es fácil escribir las dos propiedades equivalentes para funciones de densidad de variables discretas. La primera propiedad es idéntica y en la segunda propiedad sustituimos la integral por una suma sobre todos los posible valores de la variable. A la función de densidad de una variable discreta también se le conoce con los nombres de función de probabilidad o función de masa de probabilidad.



La probabilidad como un área.

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

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

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

Para el primer inciso tenemos que X es una variable aleatoria discreta que toma los valores $0, 1, 2$ y 3, con probabilidades $0, c, 2c$ y $3c$ respectivamente. Como la suma de estas probabilidades debe ser uno, obtenemos entonces 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 densidad. En el segundo inciso tenemos un ejemplo de una variable aleatoria continua que toma valores en el intervalo $[-1, 1]$. Como esta función debe integrar uno tenemos que

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

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. $\circ$

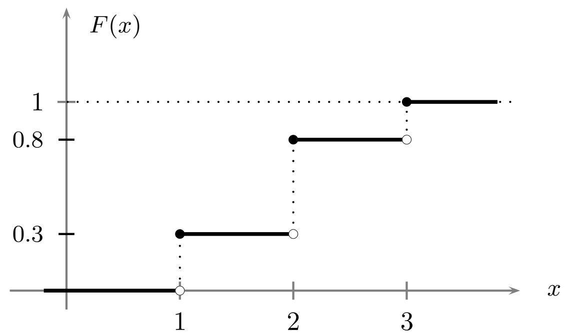
Función de distribución

La función de distribución de una variable aleatoria discreta o continua X, denotada por $F(x) : \mathbb{R} \to [0,1]$, se define como $F(x) = P(X \leq x)$. A esta importante función se le conoce también con el nombre de función de acumulación de probabilidad.

Ejemplo. (Una función de distribución discreta). Para el ejemplo anterior de la variable discreta X tenemos que la correspondiente función de distribución es

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

cuya gráfica aparece a continuación:



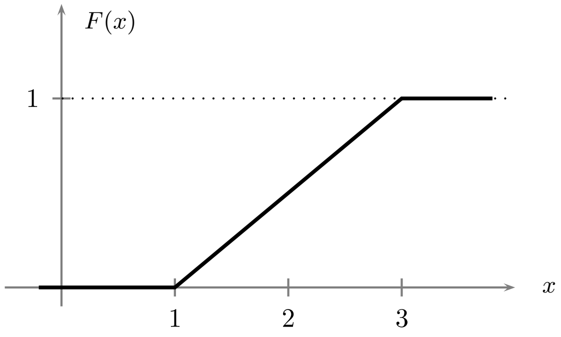
Ejemplo de una función de distribución discreta.

∘

Ejemplo. (Una función de distribución continua.) En cambio, para el ejemplo anterior de la v.a. continua X, la correspondiente función de distribución es

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

cuya gráfica aparece en la siguiente figura.



Ejemplo de una función de distribución continua.

∘

Observación. 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,

$$f(x) = \frac{d}{dx} F(x).$$

De este modo podemos encontrar $f(x)$ a partir de $F(x)$.

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 $F(x_1) \leq F(x_2)$. Esto significa que $F(x)$ es una función monótona no decreciente.

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

f) $F(x) = F(x+)$, es decir, $F(x)$ es una función continua por la derecha[2].

Demostración. Primeramente tenemos que $F(x)$ es una probabilidad pues por definición $F(x) = P(X \leq x)$. Por lo tanto se cumple la propiedad (a). 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 $\lim\limits_{x\to\infty} F(x) = P(X \leq \infty) = P(\Omega) = 1$. Análogamente el conjunto $(X \leq x)$ se aproxima al conjunto $(X \leq -\infty) = \emptyset$ cuando x tiende a menos infinito. Por lo tanto $\lim\limits_{x\to-\infty} F(x) = P(X \leq -\infty) = P(\emptyset) = 0$. Lo anterior demuestra las propiedades (b) y (c). Para demostrar (d) es suficiente observar que si $x_1 \leq x_2$ entonces $(X \leq x_1) \subseteq (X \leq x_2)$ y entonces aplicando probabilidad obtenemos $P(X \leq x_1) \leq P(X \leq x_2)$. La propiedad (f) es evidente pues 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 $P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$. Finalmente 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

$$\lim_{h\to 0} F(x+h) = F(x) + P(\emptyset) = F(x).$$

□

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 pudieramos 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 cuya función de densidad es $f(x)$, es un número denotado por $E(X)$ que se calcula como sigue

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

Ilustraremos a continuación la forma de calcular la esperanza.

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

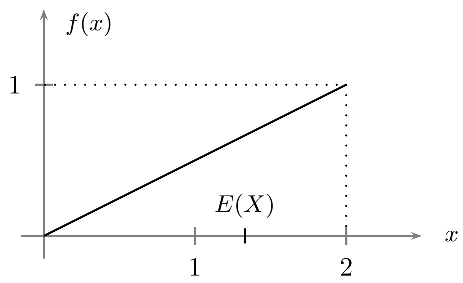
x	-1	0	1	2
$f(x)$	1/8	4/8	1/8	2/8

La esperanza de X es el número

$$E(X) = \sum_x x f(x) = (-1)(1/8) + (0)(4/8) + (1)(1/8) + (2)(2/8) = 1/2.$$

Observe que la suma su efectúa para todos los valores de x indicados en la tabla, es decir: $-1, 0, 1$ y 2. ∘

Ejemplo. Considere la variable aleatoria continua X con función de densidad $f(x) = x/2$, para $x \in (0, 2)$, cuya gráfica se muestra a continuación



La esperanza de X es entonces

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^2 x(x/2)\, dx = x^3/6 \,\Big|_0^2 = 4/3.$$

Observe que la integral sólo es relevante en el intervalo $(0, 2)$, pues fuera de dicho intervalo la función de densidad se anula. ◦

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 146 y 147. 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.

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 $Y = \exp(X^2)$ es una función de X. ¿Cuál será la esperanza de Y? El siguiente resultado es muy útil y nos dice cómo resolver este problema.

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

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

Omitiremos la demostración de la proposición anterior y nos concentraremos en su uso y aplicación.

Ejemplo. Calcularemos $E(Y)$ en donde $Y = \exp(X^2)$ y X es la variable aleatoria continua del último ejemplo. Por la proposición anterior tenemos que

$$\begin{aligned} E(Y) &= E[\exp(X^2)] \\ &= \int_{-\infty}^{\infty} \exp(x^2) f(x)\, dx \\ &= \int_0^1 \exp(x^2) \cdot 2x\, dx \\ &= e - 1. \end{aligned}$$

∘

Proposición (Propiedades de la esperanza). Sea X con esperanza finita y sea c una constante. Entonces

a) $E(c) = c$.

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

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

Demostración. Sea X la variable aleatoria constante c. Entonces por definición $E(X) = cP(X = c) = c \cdot 1 = c$. Esto demuestra el inciso (a). El inciso (b) se sigue directamente de la defnición 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 inciso (c) requiere de mayores detalles técnicos que omitiremos. □

Varianza

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

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

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). Nuevamente la correspondiente suma o integral puede no existir y en ese caso decimos que la variable aleatoria no tiene varianza finita. Observemos que para calcular $\mathrm{Var}(X)$ 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

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

∘

Ejemplo. Calcularemos la varianza de la variable aleatoria continua X con función de densidad

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

En un cálculo previo habíamos encontrado que $E(X) = 3/2$. Por lo tanto,

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} [x - E(X)]^2 f(x)\, dx \\
&= \int_0^1 \left[x - \frac{3}{2}\right]^2 \cdot 2x\, dx \\
&= \int_0^1 (2x^3 - 6x^2 + \frac{9}{2}x)\, dx \\
&= (\frac{1}{2}x^4 - 2x^3 + \frac{9}{4}x^2)\Big|_0^1 \\
&= \frac{3}{4}.
\end{aligned}$$

∘

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

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 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, y ello se muestra en el Ejercicio 156, el cual se encuentra en la página 114. □

Momentos

Finalmente definimos el *n-ésimo momento* de 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 $E[(X-\mu)^n]$, 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 continua con función de densidad $f(x)$ entonces el *n-ésimo momento* de X, si existe, se calcula como sigue:

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

Análogamente, si X es discreta con función de densidad $f(x)$, entonces

$$E(X^n) = \sum_x x^n f(x),$$

en donde la suma se efectúa sobre todos los posibles valores x que la variable aleatoria discreta X puede tomar. El *n-ésimo momento central* de X se calcula, para variables aleatorias continuas y discretas respectivamente, como indican las siguientes fórmulas:

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

1.8. Distribuciones de probabilidad

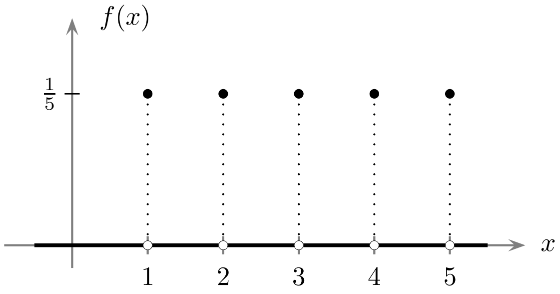
A continuación estudiaremos algunas distribuciones de probabilidad de variables aleatorias importantes. Empezaremos con las de tipo discreto y continuaremos después con las de tipo continuo. Es importante señalar que ésta es sólamente una lista parcial de algunas distribuciones de probabilidad de mayor uso.

Distribución uniforme discreta

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

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

La gráfica de la función de probabilidad de la distribución $\text{unif}\{1, 2, 3, 4, 5\}$ aparece en la siguiente figura.



Función de probabilidad de la distribución
uniforme discreta sobre el conjunto $\{1, 2, 3, 4, 5\}$.

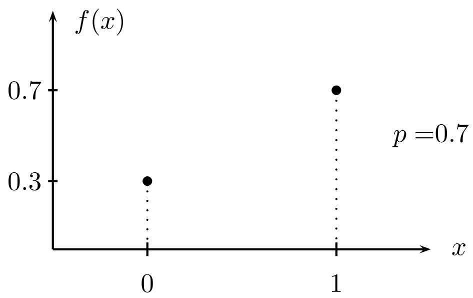
Es fácil ver que $E(X) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ y $\mathrm{Var}(X) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - E(X))^2$.

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 $X \sim \mathrm{Ber}(p)$. 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 siguiente figura.



Función de probabilidad Bernoulli.

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.

Distribución binomial

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

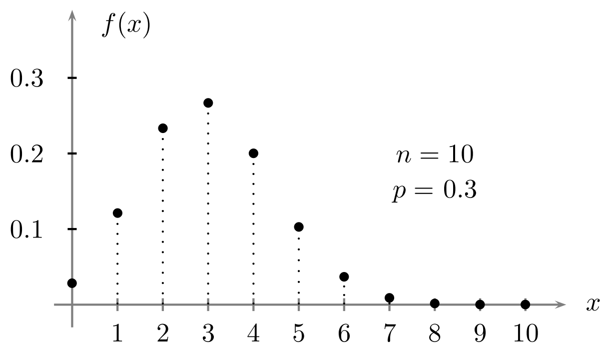
$$\Omega = \{EE\cdots EE, FE\cdots EE, \ldots, FF\cdots FF\}.$$

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

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

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

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



Función de probabilidad binomial.

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

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

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

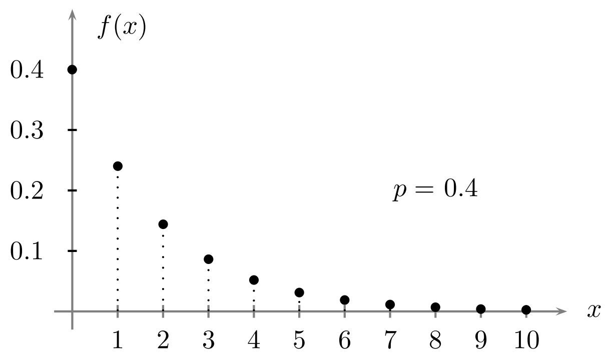
Distribución geométrica

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

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

Observamos entonces que X puede tomar los valores $0, 1, 2, \ldots$. La probabilidad de que X tome el valor x es $p(1-p)^x$. Decimos entonces que X tiene una distribución geométrica[3] con parámetro p. Y escribimos $X \sim \text{geo}(p)$ cuando

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



Función de probabilidad geométrica.

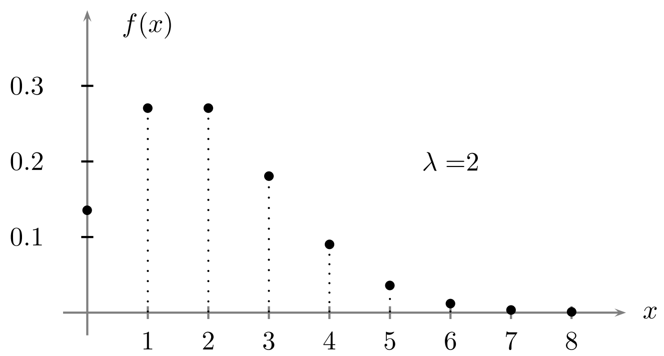
El nombre de esta distribución proviene del hecho de que cuando escribimos la suma de todas las probabilidades, obtenemos una suma geométrica. La inspección sucesiva de artículos, posiblemente para control de calidad, puede modelarse usando una distribución geométrica. Para esta distribución tenemos que $E(X) = (1-p)/p$ y $\text{Var}(X) = (1-p)/p^2$.

[3] En algunos textos definen también la distribución geométrica contando el número de ensayos (no el de fracasos) antes del primer éxito. La distribución cambia ligeramente.

Distribución Poisson

Supongamos que deseamos observar el número de ocurrencias de un cierto evento dentro de un intervalo de tiempo dado. Por ejemplo, el número de clientes que llegan a un cajero automático durante la noche, o tal vez deseamos registrar el número de accidentes que ocurren en cierta avenida durante todo un dia. Para modelar este tipo de situaciones definimos la variable aleatoria X como el número de ocurrencia de estos eventos en el intervalo de tiempo dado. Es claro entonces que X puede tomar los valores $0, 1, 2, \ldots$, y en principio no ponemos una cota superior para el número de observaciones del evento. Adicionalmente supongamos que conocemos la tasa media de ocurrencia del evento de interés, que denotamos por la letra λ (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 x se definirá como se indica a continuación. Decimos entonces que X tiene una distribución Poisson con parámetro $\lambda > 0$. Y escribimos $X \sim \text{Poisson}(\lambda)$ cuando

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



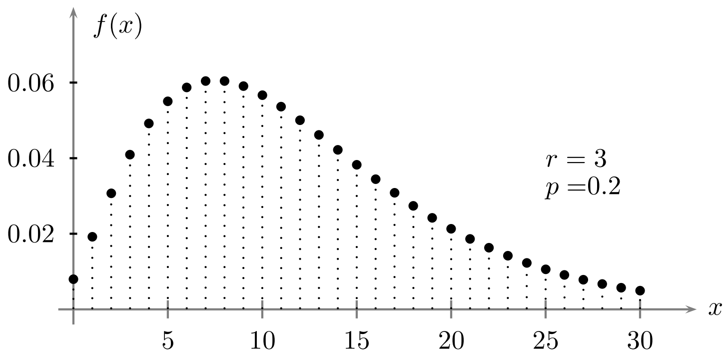
Función de probabilidad Poisson.

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

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

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

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 ahi 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.



Función de probabilidad binomial negativa.

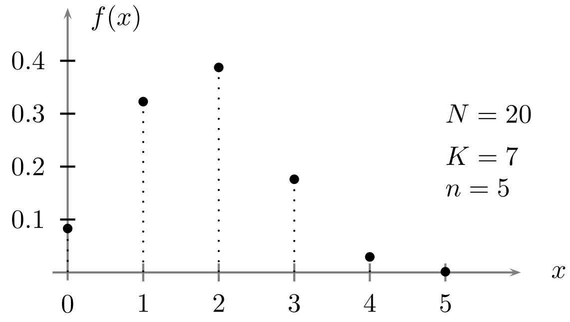
Es claro que esta distribución es una generalización de la distribución geométrica, la cual se obtiene tomando $r=1$. Además podemos demostrar que $E(X) = r(1-p)/p$ y $\text{Var}(X) = r(1-p)/p^2$.

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 entonces sin reemplazo y el orden de los objetos seleccionados no importa. El espacio muestral de este experimento consiste entonces de todas las posibles muestras de tamaño n que se pueden obtener del conjunto mayor de tamaño N. La cardinalidad del espacio muestral es entonces $\binom{N}{n}$. 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 com parámetros N, K y n. Y escribimos $X \sim \text{Hipergeo}(N, K, n)$ si

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

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



Función de probabilidad hipergeométrica.

En este caso es posible comprobar que $E(X) = nK/N$ y $\mathrm{Var}(X) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$.

Presentamos en la siguiente tabla un resumen de las distribuciones de probabilidad discretas mencionadas en este texto.

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

$f(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$
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

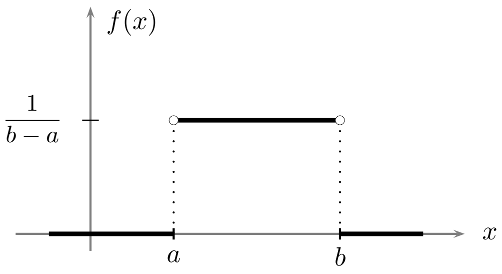
$f(x) = \binom{K}{x} \binom{N-K}{n-x} / \binom{N}{n}$
para $x = 0, 1, \ldots, n$.
Parámetros: N, K, n.
Media: nK/N.
Varianza: $nK(N-K)(N-n)/(N^2(N-1))$.

Ahora estudiaremos algunas distribuciones de probabilidad de variables aleatorias continuas.

Distribución uniforme continua

Decimos que una variable aleatoria X tiene una distribución uniforme continua en el intervalo (a, b), y escribimos $X \sim \text{unif}(a, b)$, cuando su función de densidad es

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



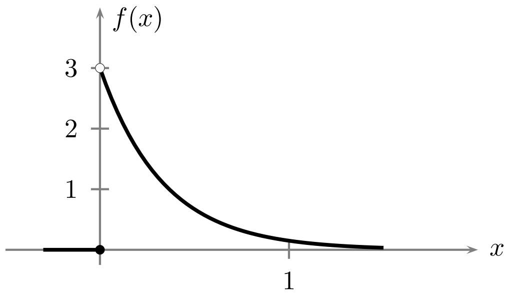
Función de densidad uniforme continua.

Es fácil verificar que $E(X) = (a + b)/2$ y $\text{Var}(X) = (b - a)^2/12$.

Distribución exponencial

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

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



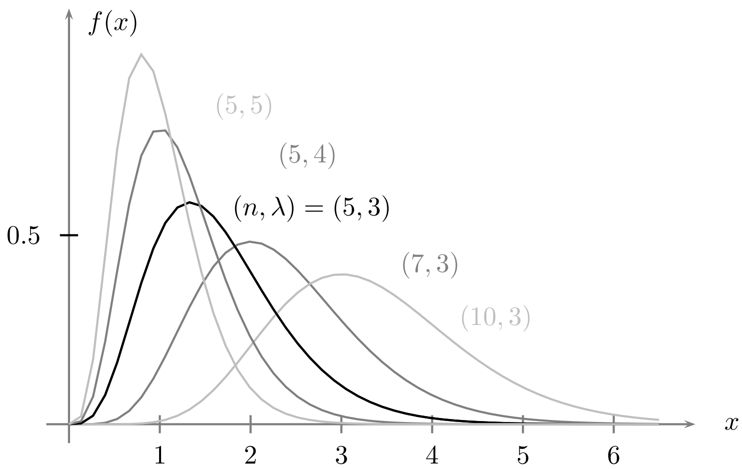
Función de densidad $\exp(\lambda)$ con $\lambda = 3$.

En este caso es muy sencillo verificar que $E(X) = 1/\lambda$ y $\text{Var}(X) = 1/\lambda^2$.

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

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



Función de densidad gama.

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

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

para valores de n para los cuales la integral es convergente. Esta función satisface 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 de probabilidad es evidente. 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 podemos demostrar que $E(X) = n/\lambda$ y $\text{Var}(X) = n/\lambda^2$.

Distribución beta

Decimos que la v.a. continua X tiene una distribución beta con parámetros $a > 0$ y $b > 0$, y escribimos $X \sim \text{beta}(a, b)$, 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 el 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 $b > 0$. No es difícil comprobar que esta función satisface la igualdad $B(a, b) = B(b, a)$, y está relacionada con la función gama a través de la identidad

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

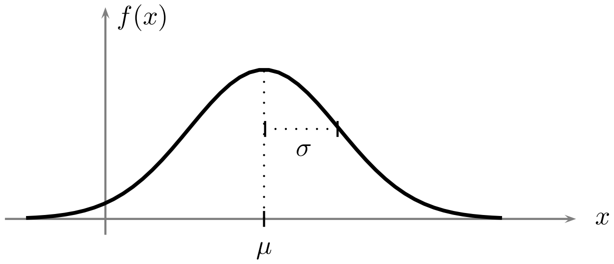
Para la distribución $\text{beta}(a, b)$ se tiene que $E(X) = a/(a+b)$ y $\text{Var}(X) = ab/[(a+b+1)(a+b)^2]$.

Distribución normal

Esta es la distribución de probabilidad de mayor importancia. Decimos que la v.a. 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 $X \sim \mathrm{N}(\mu, \sigma^2)$. La gráfica de esta función de densidad tiene forma de campana como se muestra en la siguiente figura.



Función de densidad normal o gausiana.

Observe que la campana esta centrada en el valor de μ y que se abre o cierra de acuerdo a la magnitud de σ. No es difícil probar que $E(X) = \mu$ y $\mathrm{Var}(X) = \sigma^2$. 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 sencilla

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

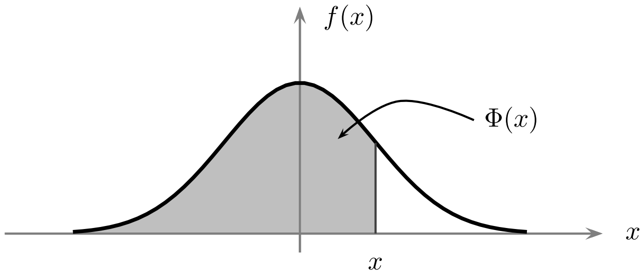
Es posible transformar una variable aleatoria normal no estándar en una estándar mediante la siguiente operación.

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

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

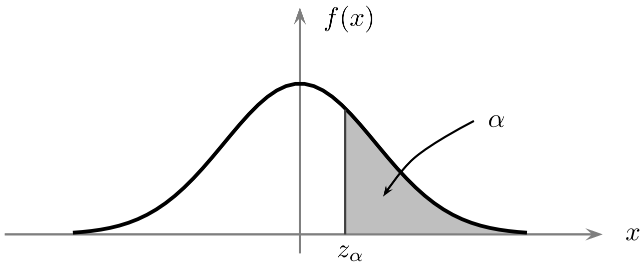
tiene una distribución normal estándar.

A la operación anterior se le conoce con el nombre de *estandarización*. Decimos también que la variable X ha sido *estandarizada*. Es común usar la letra Z para denotar una variable aleatoria con distribución normal estándar. Seguiremos nosotros también esa costumbre. En particular, para todo $x \in \mathbb{R}$, usaremos la notación $\Phi(x) = P(Z \leq x)$. Gráficamente,



Función de distribución $\Phi(x)$.

Notación z_α. Para cada $\alpha \in (0,1)$, 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, $\Phi(z_\alpha) = 1 - \alpha$. Ilustramos en la siguiente figura el significado geométrico del número z_α.



Significado gráfico 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*

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

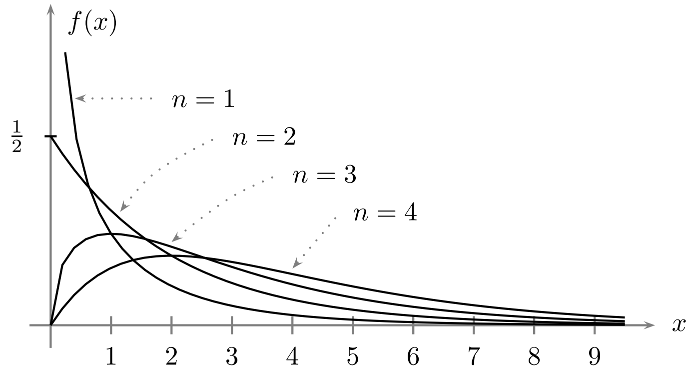
tiende a la distribución normal estándar cuando n tiende a infinito, sin importar la distribución de cada variable de la sucesiń, es decir,

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

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

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



Función de densidad ji-cuadrada.

Escribiremos simplemente $X \sim \chi^2(n)$. Podemos demostrar que $E(X) = n$ y $\mathrm{Var}(X) = 2n$. La distribución ji-cuadrada puede obtenerse del siguiente modo.

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

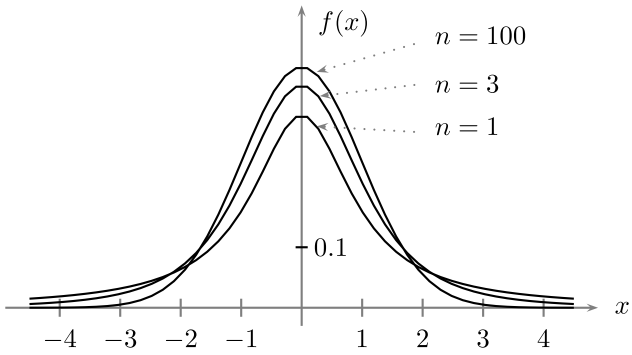
Proposición Si $X \sim \chi^2(n)$ y $Y \sim \chi^2(m)$ son dos v.a.s independientes entonces $X + Y \sim \chi^2(n+m)$.

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

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)}\,(1 + x^2/n)^{-(n+1)/2},$$

para todo $x \in \mathbb{R}$. Escribimos entonces $X \sim t(n)$.



Función de densidad t.

Es posible demostrar que $E(X) = 0$ y $\text{Var}(X) = n/(n-2)$ para $n > 2$. La distribución t se puede encontrar en los siguientes contextos.

Proposición Si $X \sim \text{N}(0,1)$ y $W \sim \chi^2(n)$ entonces

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

Proposición Sean $X_1, X_2, \ldots, X_n$ v.a.s independientes tales que $X_i \sim \text{N}(\mu, \sigma^2)$. Entonces

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

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

La siguiente tabla contiene un resumen de algunas distribuciones de probabilidad continuas.

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

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

Distribución t

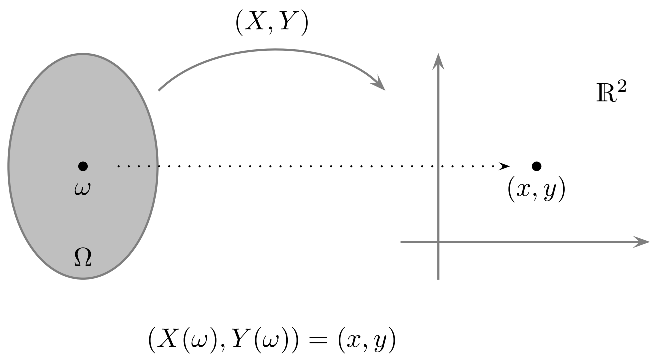
$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)}\ (1 + x^2/n)^{-(n+1)/2}$
para $x \in \mathbb{R}$.
Parámetro: $n \in \mathbb{N}$.
Media: 0.
Varianza: $n/(n-2)$.

1.9. Vectores Aleatorios

Esta sección contiene una breve introducción a las *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 (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 discretos o continuos. Un vector aleatorio (X, Y) puede considerarse como una función de Ω en $\mathbb{R}^2$ como se muestra en la siguiente figura:



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 y de densidad conjunta

Sea (X, Y) un vector aleatorio discreto que toma los valores en el conjunto $\mathrm{Ran}(X, Y) = \{(x_1, y_1), \ldots, (x_n, y_m)\}$. La función de densidad de (X, Y), denotada por $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$, se define como sigue

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

Cuando (X, Y) es un vector aleatorio continuo se dice que la función integrable y no negativa $f(x, y)$ es la función de densidad de (X, Y) si para todo $(x, y) \in \mathbb{R}^2$,

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

Toda función de densidad $f(x, y)$ satisface las siguientes dos propiedades

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 $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$ es de densidad si cumple con las dos condiciones arriba mencionadas.

Función de distribución conjunta

Sea (X, Y) un vector aleatorio discreto o continuo. La *función de distribución* de (X, Y), denotada por $F_{X,Y}(x, y) : \mathbb{R}^2 \to \mathbb{R}$, se define para cualquier par de números reales (x, y) como sigue

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

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

1. $\lim_{x,y \to \infty} F_{X,Y}(x, y) = 1.$
2. $\lim_{x \to -\infty} F_{X,Y}(x, y) = 0.$ Análogamente cuando es la variable y quien tiende a menos infinito.
3. $F_{X,Y}(x, y)$ es continua por la derecha en cada variable.
4. $F_{X,Y}(x, y)$ es una función monótona no decreciente en cada variable.
5. Para cualesquiera números $a < b$ y $c < d$ se cumple
$$F(b, d) - F(a, d) + -F(b, c) + F(a, c) \geq 0.$$

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

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

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

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

¿ Cómo encontramos $f_{X,Y}(x, y)$ a partir de $F_{X,Y}(x, y)$? Como sabemos que $f_{X,Y}(x, y)$ y $F_{X,Y}(x, y)$ guardan la relación

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

por el teorema fundamental del cálculo tenemos que

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

Función de densidad y de distribución marginal

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

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

La correspondiente función de *densidad marginal* de la variable Y se obtiene integrando respecto de la variable x, es decir,

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

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

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

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

y la correspondiente función de *distribución marginal* de la variable Y de manera análoga es la función

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

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

Independencia de variables aleatorias

Definición[Independencia de variables aleatorias] Se dice que las variables aleatorias $X_1, X_2, \ldots, X_n$ son independientes si se cumple la igualdad

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n), \tag{1.2}$$

para todo valor del vector $(x_1,\ldots,x_n)$ en $\mathbb{R}^n$.

Observe que en el lado izquierdo de la igualdad anterior aparece la función de densidad conjunta mientras que en el lado derecho se tiene el producto de las funciones de densidad marginales. En el caso particular cuando las variables aleatorias son discretas, la condición de independencia se escribe

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

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

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n). \tag{1.3}$$

Nuevamente esta igualdad debe verificarse para cada valor del vector $(x_1, \ldots, x_n)$. Las igualdades (1.2) y (1.3) son equivalentes.

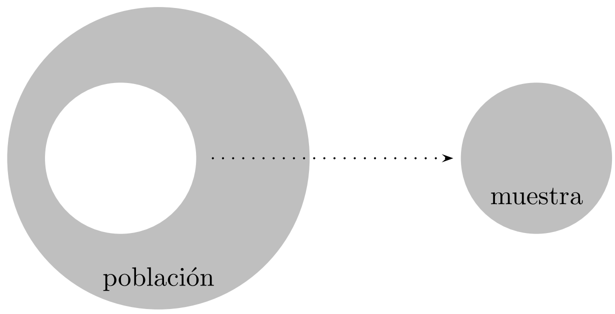
Parte 2

ESTADÍSTICA

2.1. Introducción

Población y muestra

Supondremos 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 cada elemento de la población, tomamos entonces un pequeño subconjunto de la población que llamaremos *muestra*.



Una muestra es un subconjunto de una población.

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

$$\text{Estadística} \begin{cases} \text{descriptiva,} \\ \text{inferencial.} \end{cases}$$

La *estadística descriptiva* es una colección de métodos para la organización, resumen y presentación de datos. La *estadística inferencial* consiste entonces de algunas técnicas que nos ayudan a conocer, con determinado grado de confianza, cierta información de la población con base en la información de la muestra obtenida.

2.2. Variables y tipos de datos

Una *variable* es una característica que varía de elemento a elemento en una población en estudio. Por ejemplo, si nuestra población consta de personas entonces las siguientes son ejemplos de variables que podrían interesarnos: edad, peso, sexo, estatura, etc. Las variables pueden ser *cuantitativas*, cuando se realiza una medición, o pueden ser *cualitativas*, cuando solamente presentan una cualidad. 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.

Tenemos cuatro escalas de medición para las variables, sean éstas cuantitativas o cualitativas: *escala nominal, escala ordinal, escala de intervalo y escala de razón.*

Escala nominal. La *escala nominal* está asociada a variables cualitativas y será denominada de este modo si no se pueden hacer operaciones aritméticas entre sus valores, son únicamente etiquetas. Por ejemplo, si estamos estudiando una población humana, a la variable sexo podemos asignarle dos posibles valores: F para femenino, y M para masculino, ésta es entonces una escala nominal pues los símbolos F y M son etiquetas arbitrarias, no existe orden en ellos ni podemos realizar operaciones aritméticas.

Escala ordinal. En la *escala ordinal* los valores de la variable tienen un orden pero no se pueden hacer operaciones aritméticas entre estos valores. Por ejemplo, para calificar las caracteristicas 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 de la variable en cuestión es ordinal pues existe un orden entre sus valores pero no podemos decir por ejemplo que dos valores regulares hacen un valor excelente.

Escala de intervalo. En una *escala de intervalo*, existe un orden entre los valores de la variable y existe además una noción de distancia aunque no se pueden realizar operaciones.

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

2.3. Estadística descriptiva

Supongamos que tenemos un conjunto de datos numéricos $x_1, x_2, \ldots, x_n$, que representan mediciones de alguna variable de interés. 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 y la desviación estándar.

Media

La *media* de los datos $x_1, \ldots, x_n$, denotada por $\bar{x}$, es simplemente el promedio $(x_1 + \cdots + x_n)/n$.

Moda

La *moda* es el valor observado con mayor frecuencia. La moda puede no existir para un conjunto de datos, y en caso de existir puede no ser única.

Mediana

Para calcular la *mediana* procedemos como sigue: A la muestra $x_1, x_2, \ldots, x_n$ la ordenamos de menor a mayor (incluyendo repeticiones) y obtenemos la muestra ordenada $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, 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

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

De este modo, cuando tenemos un número impar de datos, la mediana es precisamente 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 de enmedio.

Varianza y desviación estándar

La *varianza* de la muestra, 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 simplemente la raiz cuadrada positiva de s^2 y se le denota naturalmente por s.

En lo que sigue no estudiaremos muestras particulares $x_1, x_2, \ldots, x_n$ sino muestras generales $X_1, X_2, \ldots, X_n$ que pueden, en particular, tomar el conjunto de datos mencionado.

2.4. Muestras aleatorias y estadísticas

En esta sección definiremos dos términos importantes en el lenguaje de la estadística.

Muestra aleatoria

Una *muestra aleatoria* (escribimos simplemente *m.a.*) es una colección de variables aleatorias $X_1, \ldots, X_n$ para las cuales asumimos dos condiciones: independencia e idéntica distribución. De este modo, cuando se diga, por ejemplo, que una m.a. $X_1, X_2, \ldots, X_n$ es tomada de una población normal con media μ y varianza σ^2, esto significa que las variables aleatorias que forman la m.a. son todas ellas independientes entre sí y además todas ellas tienen la misma distribución normal con los mismos parámetros, es decir, $E(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2$, para $i = 1, 2 \ldots, n$. Una muestra aleatoria constituirá el elemento básico para llevar a cabo inferencias estadísticas.

Estadísticas

Una *estadística* es una función cualquiera de una muestra aleatoria. Por ejemplo, considere una muestra aleatoria dada $X_1, \ldots, X_n$. La estadística $\bar{X}$ definida como sigue

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

es efectivamente una función de la muestra aleatoria, y por lo tanto una variable aleatoria. A la estadística $\bar{X}$ se le conoce con el nombre de *media muestral*. Veamos otro ejemplo. Dada una muestra aleatoria $X_1, X_2, \ldots, X_n$, definimos la estadística *varianza muestral* como sigue

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

Estos dos ejemplos de estadísticas serán usados con frecuencia más adelante.

2.5. Estimación puntual

Denotemos por θ un parámetro desconocido de una distribución de probabilidad dada. El problema de estimación puntual consiste en encontrar un número, mediante un mecanismo lógico y con base en los datos de una muestra aleatoria, que sirva como estimación del parámetro desconocido θ.

Estimador puntual

Un *estimador puntual* para el parámetro θ, que denotaremos de manera general por $\hat{\theta}$ (se lee teta circunflejo), es una función de una muestra aleatoria $X_1, X_2, \ldots, X_n$ que sirve para estimar el valor del parámetro desconocido θ.

Insesgamiento

Un estimador puntual $\hat{\theta}$ del parámetro θ se dice que 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

Ejemplo. Sea $X_1, X_2, \ldots, X_n$ 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. Efectivamente, por la propiedad lineal de la esperanza,

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

De esta forma hemos comprobado que $\bar{X}$ es un estimador insesgado para μ. ◦

Ejemplo. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con varianza desconocida σ^2. La *varianza muestral*

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

es un estimador insesgado para la varianza desconocida σ^2. En este caso el estimador $\hat{\theta}$ es S^2 y el parámetro desconocido θ a estimar es σ^2. ◦

Método de máxima verosimilitud

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con función de densidad $f(x;\theta)$. La *función de verosimilitud* de la muestra, 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 $L(\theta)$. El valor de θ en donde se alcanza el máximo se llama *estimador de máxima verosimilitud de θ*. Ilustraremos este método con un ejemplo.

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

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

Maximizar la función $L(\mu,\sigma^2)$ es equivalente a maximizar la función $\ln L(\mu,\sigma^2)$, 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 entonces que

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

Por lo tanto

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

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

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

Observe que hemos substituido μ por $\hat\mu$ y σ^2 por $\hat\sigma^2$ pues la solución del sistema son los estimadores para μ y σ^2. De estas ecuaciones obtenemos respectivamente

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

Estos son nuestros estimadores para los parámetros μ y σ^2 de una distribución normal por el método de máxima verosimilitud. $\circ$

Método de momentos

Sea $f(x;\theta)$ la función de densidad de una variable aleatoria X que depende de un parámetro θ. Recordemos que el *k-ésimo momento* de X es el número $E(X^k)$ cuando esta esperanza existe. Ahora, dada una muestra aleatoria $X_1, X_2, \ldots, X_n$, definimos 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 consiste en igualar los momentos muestrales con los correspondientes momentos poblacionales, y resolver este sistema de ecuaciones para el parámetro θ.

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

$$\begin{array}{rrcll} & E(X) & = & \mu, & \text{(Primer momento poblacional)} \\ \text{y} & E(X^2) & = & \sigma^2+\mu^2. & \text{(Segundo momento poblacional)} \end{array}$$

El primer y segundo momento muestrales son

$$\begin{array}{rll} & \displaystyle\sum_{i=1}^{n} x_i, & \text{(Primer momento muestral)} \\ \text{y} & \displaystyle\sum_{i=1}^{n} x_i^2. & \text{(Segundo momento muestral)} \end{array}$$

La igualación respectiva produce el sistema de ecuaciones

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

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

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

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

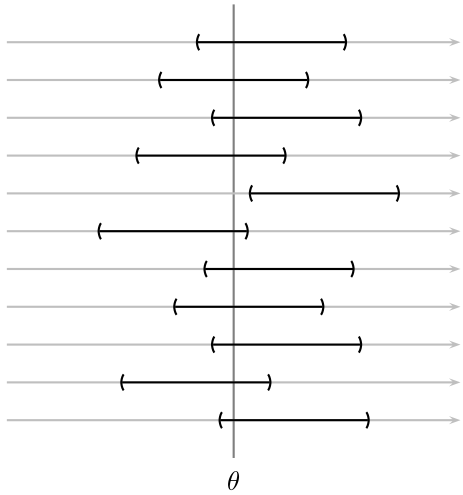
2.6. Estimación por intervalos

Alternativamente a la estimación puntual estudiada en la sección anterior, en algunos casos es preferible no dar un número como estimación 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*.

Más precisamente, un intervalo de confianza para un parámetro desconocido θ de una distribución de probabilidad dada, es un intervalo de la forma $(\hat{\theta}_1, \hat{\theta}_2)$ en donde $\hat{\theta}_1$ y $\hat{\theta}_2$ son estadísticas, esto es, funciones de una muestra aleatoria, tal que

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

en donde $\alpha \in (0, 1)$ es un número arbitrario determinado de antemano por la persona que realiza la estimación. 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, tomamos 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 $X_1, X_2, \ldots, X_n$, de modo que al tomar estas variables aleatorias distintos valores se generan distintos intervalos de confianza como se muestra a continuación.



Diferentes valores del intervalo aleatorio $(\hat{\theta}_1, \hat{\theta}_2)$,
algunos de ellos conteniendo el verdadero valor de θ.

De esta forma podemos decir que el intervalo aleatorio $(\hat{\theta}_1, \hat{\theta}_2)$ contiene el verdadero valor del parámetro θ con probabilidad $1-\alpha$. 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 para algunos casos particulares.

Intervalo para la media de una población normal con varianza conocida

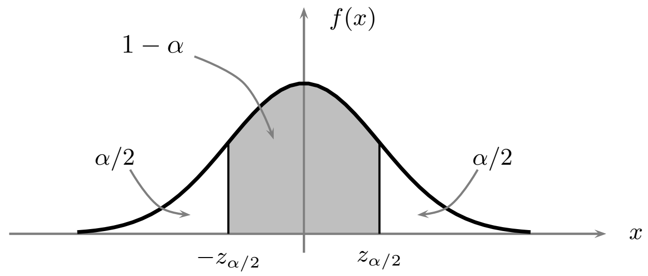
Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población normal con media *desconocida* μ y varianza *conocida* σ^2. Ilustraremos a continuación una forma de encontrar un intervalo de confianza al $(1-\alpha)100\,\%$ para el parámetro desconocido μ. Como cada una de las variables de la muestra tiene distribución $\mathrm{N}(\mu, \sigma^2)$, la variable $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 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 tal que

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

Gráficamente,



Distribución normal estándar

Despejando la constante desconocida μ, obtenemos

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

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 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 $x_1, x_2, \ldots, x_{20}$ arrojan una media muestral $\bar{x}$ de 1050 horas. Si consideramos un nivel de confianza del 95 %, es decir, α =0.05, entonces de la tabla de probabilidad normal se encuentra que $z_{\alpha/2}$ =1.65, y entonces puede ahora calcularse el intervalo

$$\begin{aligned}(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) &= (1050 - 1{,}65 \cdot \frac{30}{\sqrt{20}},\ 1050 + 1{,}65 \cdot \frac{30}{\sqrt{20}}) \\ &= (1038{,}93,\ 1061{,}06).\end{aligned}$$

De esta forma el intervalo (1038.93,1061.06) contiene el verdadero valor de la vida promedio útil de todos los focos producidos por la compañía con confianza del 95 %. ∘

Intervalo para la media de una población normal con varianza desconocida

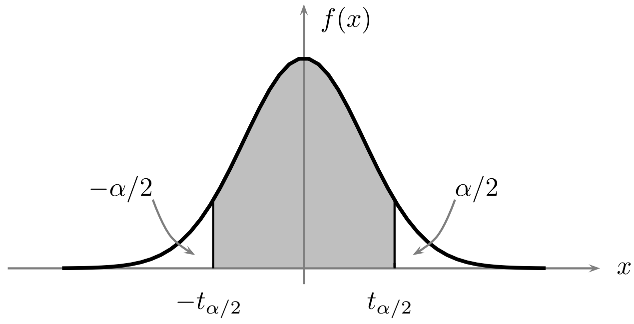
Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población normal con media *desconocida* μ y varianza *desconocida* σ^2. Tenemos entonces que la variable aleatoria

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

tiene una distribución t con $n-1$ grados de libertad. Observe que ésta 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 la forma siguiente. 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 tal que

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

Gráficamente,



Distribución t con $n-1$ grados de libertad.

Despejando la constante desconocida μ, obtenemos

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

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 *exacto* al $(1-\alpha)100\,\%$ para la media desconocida μ de una población normal, para cualquier tamaño de muestra y sin suponer la varianza conocida. No olvidemos que el valor $t_{\alpha/2}$ corresponde a la distribución t con $n-1$ grados de libertad.

Intervalo *aproximado* para la media de una población con varianza desconocida y tamaño de muestra grande

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población normal con media *desconocida* μ y varianza *desconocida* σ^2. Supongamos que el tamaño n de la muestra es grande, i.e. $n \geq 30$. Entonces la variable aleatoria

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

tiene una distribución *aproximada* normal estándar. Esto es una consecuencia del teorema del límite central pues el tamaño de la muestra es grande. En este caso también podemos encontrar un intervalo *aproximado* de confianza para el parámetro desconocido μ. El procedimiento es análogo al anterior. 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 la constante desconocida μ, obtenemos

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

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. Ilustremos lo anterior con un ejemplo.

A manera de resumen,

Hipótesis	Intervalo para la media μ de una pobación normal
Varianza σ^2 conocida Cualquier tamaño de muestra	$P(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha.$
Intervalo aproximado Varianza σ^2 desconocida Muestra grande, $n \geq 30$	$P(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}) \doteq 1-\alpha.$
Varianza σ^2 desconocida Cualquier tamaño de muestra	$P(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}) = 1-\alpha.$

2.7. Pruebas de hipótesis

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 bin(n,p) entonces la afirmación "$p = 0{,}2$" es una hipótesis. Si X tiene una distribución N(μ, σ^2) entonces la afirmación "$\mu > 0$" es otro ejemplo de hipótesis estadística.

Un hipótesis es *simple* si especifica por completo la distribución de probabilidad en cuestión. Por ejemplo, si X tiene una distribución exp(λ) entonces la afirmación "$\lambda = 5$" es una hipótesis simple. Si X tiene una distribución 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
compuesta	vs	compuesta

Ahora podemos entonces definir una *prueba de hipótesis* como *una regla para decidir* si aceptamos la hipótesis nula (H_0) o la rechazamos en favor de la hipótesis alternativa (H_1). Al tomar una decisión de este tipo podemos cometer errores sin saberlo. 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*, a la probabilidad de cometer este segundo tipo de error se le denota por la letra β. Tenemos entonces la siguiente tabla

	H_0 cierta	H_0 falsa
Rechazar H_0	Error tipo I	Decisión correcta
No rechazar H_0	Decisión correcta	Error tipo II

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

Prueba de hipótesis acerca de la media de una población 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

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

Queremos contrastar las hipótesis

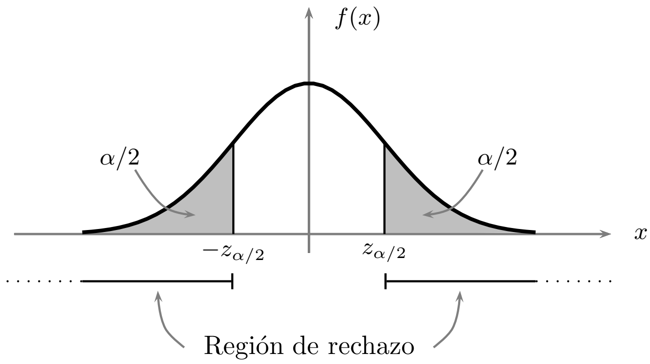
$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0.$$

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

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

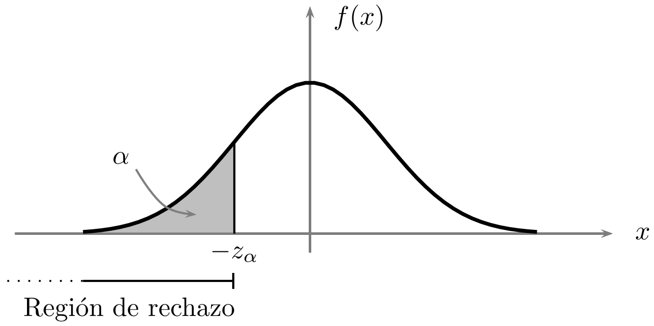
La estadística $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ es una medida natural de la distancia entre $\bar{X}$ (el estimador de μ) y su valor esperado μ_0 (cuando H_0 es cierta). Es entonces natural rechazar la hipótesis nula H_0 cuando la variable Z sea grande. Es por esto que tomamos como criterio de decisión rechazar la hipótesis nula $H_0 : \mu = \mu_0$ 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 $P(|Z| \geq z_{\alpha/2}) = \alpha$. Este valor $z_{\alpha/2}$ es precisamente la constante k pues con esto garantizamos que la región de rechazo sea de tamaño α. En resumen,

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

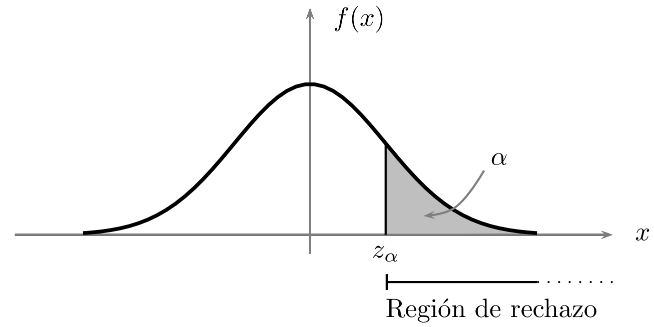


Otras pruebas son:

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



Prueba: $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu > \mu_0$.
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: $\Phi\left(z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right)$, para $\mu_1 > \mu_0$.



Estas tres pruebas y sus correspondientes regiones de rechazo tienen un nivel de significancia de tamaño α, esto es, la probabilidad de cometer el error tipo I (rechazar H_0 cuando ésta es verdadera) es $\alpha \in (0, 1)$. El investigador puede asignar o establecer a priori esta probabilidad α, pero ¿cómo se calculan las probabilidades de cometer el error tipo II? Esta probabilidad se denota por la letra griega β

y la hemos definido como la probabilidad de no rechazar la hipótesis nula H_0 cuando ésta es falsa. Calcularemos a continuación esta probabilidad para la prueba $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu > \mu_0$. Sea μ_1 cualquier número tal que $\mu_1 > \mu_0$, entonces

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

De manera análoga se calcula esta probabilidad en los otros dos casos.

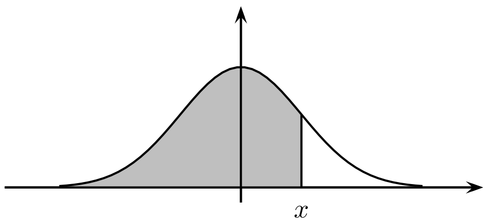
Apéndice A

Formulario

A.1. 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

A.2. 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

Apéndice B

Ejercicios

Conjuntos

1. ¿Qué es la teoría de la probabilidad?
2. ¿Qué es un experimento aleatorio?
3. ¿Qué es el espacio muestral de un experimento aleatorio?
4. Determine el espacio muestral del experimento aleatorio consistente en lanzar una moneda tres veces.
5. Determine el espacio muestral del experimento aleatorio consistente en lanzar a un mismo tiempo tres monedas indistinguibles.
6. Determine el espacio muestral del experimento aleatorio consistente en escoger un número real al azar dentro del intervalo $[-1, 1]$ y después elevarlo al cuadrado.
7. Determine el espacio muestral del experimento aleatorio consistente en registrar el número de llamadas telefónicas que llegan a un conmutador en un minuto dado.

8. Determine el espacio muestral del experimento aleatorio consistente en colocar al azar dos bolas distinguibles en cuatro celdas numeradas.

9. Determine el espacio muestral del experimento aleatorio consistente en colocar al azar dos bolas indistinguibles en cuatro celdas numeradas.

10. Determine el espacio muestral del experimento aleatorio consistente en observar el marcador final de un juego de futbol soccer.

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

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

    a) $\Omega = \{0, 1, 2, 3, \ldots\}$.
    b) $\Omega = (0, \infty)$.

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

    a) $\Omega = \{2, 4, 6, \ldots\}$.
    b) $\Omega = [1, \infty)$.
    c) $\Omega = \{0, 1\}$.

14. Determine un experimento aleatorio en el que el espacio de posibles resultados sea el conjunto de números racionales.

15. Determine un experimento aleatorio en el que el espacio de posibles resultados sea el conjunto de números complejos.

16. ¿Qué es un evento?

17. Demuestre las leyes distributivas:

    a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
    b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

18. Demuestre las leyes de De Morgan:

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

19. Enuncie y demuestre las leyes de De Morgan para tres conjuntos. Para la demostración use la validez del mismo resultado para dos conjuntos.

20. Demuestre que $A = (A \cap B) \cup (A \cap B^c)$.

21. Demuestre que $A \cap B = B \cap (A \cup B^c)$.

22. Demuestre que $A \cap B = A \cap (B \cup A^c)$.

23. Demuestre que $A \cup B = B \cup (A \cap B^c)$.

24. Demuestre que $A \cup B = A \cup (B \cap A^c)$.

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

25. Demuestre que

    a) $A - B = A - (A \cap B)$.
    b) $A - B = (A \cup B) - B$.

26. Demuestre que

    a) $A \cap (B - C) = (A \cap B) - (A \cap C)$.
    b) $(A - C) \cap (B - C) = (A \cap B) - C$.

27. Demuestre que las siguientes dos definiciones de la operación diferencia simétrica son equivalentes.

    a) $A \triangle B = (A - B) \cup (B - A)$.
    b) $A \triangle B = (A \cup B) - (A \cap B)$.

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

29. 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 $A \triangle B$.

30. 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 $A \triangle B$.

31. 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 $A \triangle B$.

32. 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 $A \triangle B$.

Probabilidad

33. Enuncie con precisión los tres axiomas de la probabilidad y cuatro propiedades que se pueden demostrar usando los axiomas.

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

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

36. 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.

37. Demuestre que $P(A^c) = 1 - P(A)$.

38. Demuestre que $P(\emptyset) = 0$,

    a) usando $P(\Omega) = 1$.
    b) sin usar $P(\Omega) = 1$.

39. Demuestre que si $A \subseteq B$, entonces $P(A) \leq P(B)$.

40. Demuestre que $0 \leq P(A) \leq 1$.

41. Demuestre que para cualesquiera dos eventos A y B,

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

42. Demuestre que si $A \subseteq B$, entonces $P(B - A) = P(B) - P(A)$.

43. Demuestre que para cualesquiera dos eventos A y B,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

44. Demuestre que para cualesquiera tres eventos A, B y C,
$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad -P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad +P(A \cap B \cap C). \end{aligned}$$

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

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

47. 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.

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

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

50. 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 $A \subseteq B$.

51. 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 $P(A^c) > 0$.

52. 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 $P(A^c) < 1/2$.

Análisis combinatorio

53. Enuncie con precisión el principio de multiplicación.
54. ¿Qué es una ordenación con repetición?
55. ¿Qué es una permutación de n objetos?
56. ¿De cuántas formas distintas pueden seis personas formarse en una fila lineal?
57. 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 $1/60$.
58. ¿Qué es una permutación de n en k?
59. ¿Cuántas diagonales se pueden trazar en un polígono convexo de n lados?
60. ¿De cuántas maneras diferentes pueden clasificarse los tres primeros lugares de una carrera de 15 corredores ?
61. ¿Cuántos enteros positivos de a lo mas cinco dígitos son divisibles por 2? ¿Cuántos hay que empiecen con el dígito 1?
62. ¿Qué es una combinación de n en k?
63. ¿Cuántos equipos de 3 personas se pueden formar de un grupo de 5 personas?
64. Demuestre que $\left(\begin{array}{c} n \\ k \end{array} \right) = \dfrac{n-k+1}{k} \left(\begin{array}{c} n \\ k-1 \end{array} \right)$.
65. Demuestre que $\left(\begin{array}{c} n \\ k \end{array} \right) = \dfrac{n}{n-k} \left(\begin{array}{c} n-1 \\ k \end{array} \right)$.
66. Demuestre que $\left(\begin{array}{c} n-1 \\ k \end{array} \right) = \dfrac{k+1}{n} \left(\begin{array}{c} n \\ k+1 \end{array} \right)$.
67. Demuestre que $\left(\begin{array}{c} n \\ k \end{array} \right) = \left(\begin{array}{c} n-1 \\ k \end{array} \right) + \left(\begin{array}{c} n-1 \\ k-1 \end{array} \right)$. Esta es la fórmula para construir el triángulo de Pascal.
68. 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? ($0 \leq k \leq 20$).

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

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

71. *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.

72. 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?

73. 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?

74. ¿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.

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

76. 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 llevaran 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?

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

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

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

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

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

82. 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?

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

84. ¿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?

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

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

86. 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?

87. ¿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?

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

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

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

91. 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) $x_1 + x_2 + \cdots + x_k \geq n$?

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

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

93. Desarrolle la expresión $(a+b+c)^3$ usando la fórmula de coeficientes multinomiales (1.1) en la página 25, y después compruebe la fórmula directamente multiplicando el trinomio por si mismo.

Probabilidad condicional e independencia

94. Enuncie la definición de probabilidad condicional e indique una interpretación de ella.

95. 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.

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

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

97. 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.

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

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

100. 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) $P(A|A \cap B) = P(B|A \cap B) = 1$.

101. Diga falso o verdadero. Justifique en cada caso.
    a) $P(\Omega|B) = 1$.
    b) $P(\emptyset|B) = 0$.
    c) $P(A|A) = P(A)$.

102. 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) $P(A_1 \cap A_2|B) = P(A_1|B)P(A_2|B)$.

103. Enuncie con precisión y demuestre la regla de multiplicación.

104. ¿Cuándo decimos que dos eventos A y B son independientes?

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

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

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

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

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

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

109. Sean A y B eventos independientes. Demuestre que

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

110. 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.

111. 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.

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

113. 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.

114. ¿Cuándo decimos que n eventos $A_1, A_2, \ldots, A_n$ son independientes?

115. ¿Cuántas igualdades son necesarias verificar para demostrar que n eventos son independientes?

116. Enuncie con precisión y demuestre el teorema de probabilidad total.

117. *La urna de Polya.* Suponga que en una urna se tienen b bolas blancas y r bolas rojas. El experimento aleatorio consiste en seleccionar una bola al azar y regresarla a la urna junto con c bolas del mismo color. Sea el evento R_n =“Se seleciona una bola roja en la n-ésima extracción”. Claramente $P(R_1) = r/(r+b)$.

    a) Demuestre que $P(R_2) = \dfrac{r}{r+b}$.

    b) Ahora use el método de inducción para demostrar que $P(R_n) = r/(r+b)$ para cualquier $n \geq 3$. Sorprendentemente la respuesta no depende de n.

118. Una persona toma al azar uno de los números $1, 2$ o 3 y luego tira un dado 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?

119. Enuncie con precisión y demuestre el teorema de Bayes.

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

121. Se codifica un mensaje en sistema binario y se envía a través de un canal de transmisión. La probabilidad de transmisión de un “0” es 0.4 y la probabilidad de transmisión de un “1” es 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. Se envía un dígito escogido al azar, 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”.

Funciones de densidad y de distribución

122. Enuncie la definición de función de densidad para una v.a. tanto en el caso discreto como en el continuo. Enuncie además las dos propiedades que la caracterizan.

123. Encuentre el valor de la constante c para que la siguiente función sea una función de densidad. Grafique $f(x)$ y calcule $P(X \in \{2,3,4\})$.

$$f(x) = \begin{cases} cx & \text{si } x = 1, 2, \ldots, 10. \\ 0 & \text{otro caso.} \end{cases}$$

124. Encuentre el valor de la constante c para que la siguiente función sea de densidad.

$$f(x) = \begin{cases} cx^2 & \text{si } x = 1, 2, \ldots, 10. \\ 0 & \text{otro caso.} \end{cases}$$

125. 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 $P(X \in [-\pi, 2\pi])$.

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

126. Encuentre el valor de la constante c para que la siguiente función sea de densidad. Grafique $f(x)$ y calcule $P(X \in (1, \infty))$.

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

127. Explique porqué no es posible encontrar un valor de la constante c para que la siguiente función sea de densidad.

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

128. Explique porqué no es posible encontrar un valor de la constante c para que la siguiente función sea de densidad.

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

129. 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

130. Sea X discreta con función de densidad dada por la siguiente tabla.

x	-2	-1	0	2	3	5
$f(x)$	0.1	0.15	0.4	0.1	0.15	0.1

a) Grafique $f(x)$.

b) Calcule la función de densidad de las siguientes v.a.s $Y = X^2$, $Z = |X|$ y $W = 2X - 5$. Grafique en cada caso.

131. Sea X discreta con función de densidad dada por la siguiente tabla

x	-2	0	2
$f(x)$	0.1	c	0.1

a) Encuentre el valor de c.

b) Grafique $f(x)$.

c) Calcule y grafique la función de probabilidad de la variable $Y = X^2$.

132. 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 $P(|X - 1| \geq m) = 1/2$.

133. Sea X una v.a. con función de distribución

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

Grafique $F(x)$. Encuentre y grafique la correspondiente función de densidad $f(x)$.

134. Sea X una v.a. con función de distribución

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

Grafique $F(x)$. Encuentre y grafique la correspondiente función de densidad $f(x)$.

135. 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 v.a. que denota la suma de los números de las dos bolas seleccionadas.

a) Determine Ω.

b) Determine y grafique $f(x)$.

c) Calcule y grafique $F(x)$.

d) Calcule $P(X \geq 6)$, $P(3 < X \leq 5)$ y $P(X = 6)$.

136. Determine si la siguiente función es de densidad. Grafique la función y justifique su respuesta.

$$f(x) = \begin{cases} 1/6 & \text{si } x = 0, 1, \\ 2/3 & \text{si } x = 2, \\ 0 & \text{otro caso.} \end{cases}$$

137. Determine si la siguiente función es de densidad. Grafique la función y justifique su respuesta.

$$f(x) = \begin{cases} \left(\begin{array}{c} 4 \\ x \end{array} \right) \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{4-x} & \text{si } x = 0, 1, 2, 3, 4, \\ 0 & \text{otro caso.} \end{cases}$$

138. Determine si la siguiente función es de densidad. Grafique la función y justifique su respuesta.

$$f(x) = \begin{cases} \frac{4}{5}x & \text{si } x \in [0, 2], \\ 0 & \text{otro caso.} \end{cases}$$

139. Determine si la siguiente función es de densidad. Grafique la función y justifique su respuesta.

$$f(x) = \begin{cases} \frac{2}{3}x^2 - 2x + \frac{4}{3} & \text{si } x \in [0, 3], \\ 0 & \text{otro caso.} \end{cases}$$

140. Sea X una v.a. con función de distribución

$$F(x) = \begin{cases} 1 - \left(\frac{1}{2}\right)^{x+1} & \text{si } x = 0, 1, 2, 3, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

Encuentre y grafique $f(x)$. Calcule $P(0 \leq X < 10)$.

141. Sea X una v.a. continua con función de densidad

$$f(x) = \begin{cases} \frac{2}{9}x & \text{si } 0 \leq x \leq c, \\ 0 & \text{otro caso.} \end{cases}$$

Encuentre el valor de la constante c. Encuentre y grafique $F(x)$.

Esperanza, varianza, momentos

142. Escriba la definición de esperanza de una variable aleatoria tanto en el caso discreto como en el caso continuo y mencione una interpretación de ella.

143. Sea a un número real fijo. Construya una variable aleatoria X tal que $E(X) = a$.

144. Calcule la esperanza de la variable aleatoria discreta X cuya función de densidad es

    a)

145. Calcule la esperanza de la variable aleatoria continua X cuya función de densidad es

    a) $f(x) =$.

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

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

    Demuestre que $f(x)$ es efectivamente una función de densidad y que $E(X)$ no existe. Este es un ejemplo de una variable aleatoria discreta que no tiene esperanza finita.

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

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

    Compruebe que $E(X)$ no existe. Este es un ejemplo de una variable aleatoria continua que no tiene esperanza finita.

148. 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.

149. Demuestre que

a) $E(E(X)) = E(X)$.

b) $\text{Var}(\text{Var}(X)) = 0$.

150. Sea X la variable aleatoria constante c. Use la definición de esperanza y varianza para demostrar que

a) $E(X) = c$.

b) $E(X^n) = c^n$.

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

151. Calcule la media y varianza de la variable aleatoria X con función de densidad

$$f(x) = \begin{cases} 1/9 & \text{si } x = 0, 1, 2, \\ 2/9 & \text{si } x = 3, 4, 5, \\ 0 & \text{otro caso.} \end{cases}$$

152. Calcule la media y varianza de la variable aleatoria X cuya función de densidad es

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

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

a) $\text{Var}(E(X)) = 0$.

b) $E(\text{Var}(X)) = \text{Var}(X)$.

154. Sea X una variable aleatoria continua con función de densidad $f(x) = \frac{1}{2}e^{-|x|}$. 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 $n = 0, 1, 2, \ldots$

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

a) $E(-X) = -E(X)$.

b) $\text{Var}(-X) = -\text{Var}(X)$.

c) $E(\text{Var}(X)) = \text{Var}(E(X))$.

156. Sean X y Y dos variables aleatorias con funciones de densidad dadas por las tablas siguientes.

x	-1	0
$f_X(x)$	$1/2$	$1/2$

y	0	1
$f_Y(y)$	$1/2$	$1/2$

Demuestre que $\text{Var}(X+Y) \neq \text{Var}(X) + \text{Var}(Y)$.

157. Encuentre el error en la siguiente “demostración” de la afirmación de que la varianza de cualquier variable aleatoria es cero.

$$\begin{aligned} 0 &= \text{Var}(0) \\ &= \text{Var}(X + (-X)) \\ &= \text{Var}(X) + \text{Var}(-X) \\ &= \text{Var}(X) + \text{Var}(X) \\ &= 2\text{Var}(X). \end{aligned}$$

Distribuciones de probabilidad

158. Sea X con distribución uniforme en el conjunto $\{1, \ldots, n\}$. Demuestre que

a) $E(X) = \dfrac{(n+1)}{2}$.

b) $E(X^2) = \dfrac{(n+1)(2n+1)}{6}$.

c) $\text{Var}(X) = \dfrac{(n^2-1)}{12}$.

159. 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?

160. Sea X una variable aleatoria con distribución $\text{Ber}(p)$. Verifique que $f(x)$ es efectivamente una función de densidad, y demuestre que

a) $E(X) = p$.

b) $E(X^n) = p$, para $n \geq 1$.

c) $\text{Var}(X) = p(1-p)$.

161. Sea X una variable aleatoria con distribución bin(n, p). Verifique que $f(x)$ es efectivamente una función de densidad.

162. Sea X una v.a. con distribución bin(n, p). Demuestre que

a) $E(X) = np$.

b) $\text{Var}(X) = np(1-p)$.

163. 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?

164. Sea X una variable aleatoria con distribución bin(n, p). Demuestre que la variable $Y = n - X$ tiene una distribución bin$(n, 1-p)$. Proporcione una explicación probabilísta de este resultado.

165. Sea X con distribución bin(n, p). Demuestre que para $x = 0, 1, \ldots, n-1$,

$$P(X = x+1) = \frac{p}{1-p} \cdot \frac{n-x}{x+1} \cdot P(X = x).$$

166. Se lanza una moneda equilibrada 6 veces. Calcule la probabilidad de que cada cara caiga exactamente 3 veces.

167. Se lanza una moneda equilibrada $2n$ veces. Calcule la probabilidad de que ambas caras caigan el mismo número de veces.

168. Sea X una variable aleatoria con distribución bin(n, p). Demuestre que $0 \leq \text{Var}(X) \leq E(X)$.

169. Sea X una variable aleatoria con distribución Poisson(λ). Verifique que $f(x)$ es efectivamente una función de densidad y demuestre que

a) $E(X) = \lambda$.

b) $\text{Var}(X) = \lambda$.

170. Sea X una v.a. con distribución Poisson(λ). Demuestre que para $x = 0, 1, 2, \ldots$

$$P(X = x+1) = \frac{\lambda}{x+1} \cdot P(X = x).$$

171. Sea X una variable aleatoria con distribución Poisson(λ). Demuestre que la probabilidad de que X tome un valor par es $(1+e^{-2\lambda})/2$.

172. 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.

173. En promedio se reciben 2 peticiones de acceso a una página web durante un minuto cualquiera. Utilice el modelo Poisson para calcular la probabilidad de que en un minuto dado cualquiera,

    a) nadie solicite acceso a la página.

    b) se reciban mas de dos peticiones.

174. El número de aviones que llegan a un aeropuerto internacional tiene una distribución Poisson con una frecuencia de 5 aviones cada 10 minutos. Calcule la probabilidad de que

    a) no llegue ningún avión en un periodo de 20 minutos.

    b) no llegue ningún avión en el minuto siguiente.

    c) llegue solo un avión en un periodo de 20 minutos.

175. 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) ¿Cuál es el número de computadoras con falla más probable en un mes?

    c) Responda a los incisos anteriores si reducimos la capacidad de reparación del laboratorio a una computadora por mes.

176. Sea X una v.a. con distribución geo(p). Verifique que $f_X(x)$ es efectivamente una función de densidad.

177. Sea X una v.a. con distribución geo(p). Demuestre que

    a) $E(X) = \dfrac{1-p}{p}$.

b) $\text{Var}(X) = \dfrac{1-p}{p^2}$.

178. Sea X una v.a. con distribución geo(p). Demuestre que para $a, b = 0, 1, 2, \ldots$,

$$P(X \geq a + b \,|\, X \geq a) = P(X \geq b).$$

179. Sea X una v.a. con distribución bin neg(p, r). Verifique que $f_X(x)$ es efectivamente una función de densidad.

180. Sea X una v.a. con distribución bin neg(p, r). Demuestre que

a) $E(X) = r\left(\dfrac{1-p}{p}\right)$.

b) $\text{Var}(X) = r\left(\dfrac{1-p}{p^2}\right)$.

181. Sea X una v.a. con distribución unif$[a, b]$. Verifique que $f_X(x)$ es efectivamente una función de densidad.

182. Sea X una v.a. con distribución uniforme en el intervalo $[a, b]$. Demuestre que

a) $E(X) = \dfrac{a+b}{2}$.

b) $\text{Var}(X) = \dfrac{(b-a)^2}{12}$.

183. Sea X una v.a. con distribución uniforme en el intervalo $[a, b]$. Demuestre que

$$E(X^n) = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}.$$

184. Sea X una v.a. con distribución uniforme en el intervalo $[0, 1]$. Demuestre que

$$E(X^n) = \frac{1}{n+1}.$$

185. Sea X una v.a. con distribución uniforme en el intervalo $[-1, 1]$. Demuestre que

$$E(X^n) = \begin{cases} \dfrac{1}{n+1} & \text{si } n \text{ es par,} \\ 0 & \text{si } n \text{ es impar.} \end{cases}$$

186. Sea X una v.a. con distribución uniforme en el intervalo $[0,1]$. Obtenga la distribución de la v.a. $Y = 10X - 5$.

187. Sea X una v.a. con distribución uniforme en el intervalo $[0,1]$. Obtenga la distribución de la v.a. $Y = 4X(1-X)$.

188. Sea X una v.a. con distribución $\exp(\lambda)$. Verifique que $f_X(x)$ es efectivamente una función de densidad.

189. Sea X una v.a. con distribución $\exp(\lambda)$. Demuestre que

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

190. Sea X una v.a. con distribución $\exp(\lambda)$. Demuestre que

    a) $E(X) = 1/\lambda$.

    b) $\mathrm{Var}(X) = 1/\lambda^2$.

191. Sea X una v.a. con distribución $\exp(\lambda)$. Demuestre que para cualesquiera $x, y \geq 0$,

$$P(X \geq x + y \,|\, X \geq x) = P(X \geq y).$$

La distribución exponencial es la única distribución continua que satisface esta propiedad llamada *pérdida de memoria*.

192. Sea X una v.a. con distribución $\exp(\lambda)$. Demuestre que para todo $x, y \geq 0$,

$$F_X(x+y) - F_X(y) = F_X(x)[1 - F_X(y)].$$

193. Suponga que el tiempo que un usuario cualquiera permanece conectado a un servidor en una red se puede modelar como una v.a. con distribución exponencial con media igual a 10 minutos. Calcule 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.

¿De mil usuarios, cuántos tienen un conexión superior a una hora?

194. Sea X una v.a. con distribución $\mathrm{gama}(\lambda, n)$. Verifique que $f_X(x)$ es efectivamente una función de densidad.

195. Sea X una v.a. con distribución $\mathrm{gama}(\lambda, n)$. Demuestre que

a) $E(X) = n/\lambda$.

b) $\mathrm{Var}(X) = n/\lambda^2$.

c) $E(X^m) = \frac{\Gamma(M+n)}{\lambda^m \Gamma(n)}$.

196. Demuestre las siguientes propiedades de la función gama.

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

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

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

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

197. Sea X una v.a. con distribución beta(a, b). Verifique que $f_X(x)$ es efectivamente una función de densidad.

198. Sea X una v.a. con distribución beta(a, b). Demuestre que

a) $E(X) = \dfrac{a}{a+b}$.

b) $\mathrm{Var}(X) = \dfrac{ab}{(a+b+1)(a+b)^2}$.

199. Demuestre las siguientes propiedades de la función beta.

a) $B(a, b) = B(b, a)$.

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

c) $B(a, 1) = 1/a$.

d) $B(1, b) = 1/b$.

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

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

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

h) $B(1/2, 1/2) = \pi$.

200. Sea X una v.a. con distribución beta$(1/2, 1/2)$. En este caso decimos que X tiene una *distribución arcoseno*.

a) Calcule y grafique $f_X(x)$.

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

c) Calcule directamente $E(X)$ y $\mathrm{Var}(X)$.

201. Sea X una v.a. con distribución beta(a,b). Demuestre que cuando $a>0$ y $b=1$, la función de distribución de X toma la forma

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ x^a & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

202. Sea X una v.a. con distribución beta(a,b). Demuestre que cuando $a=1$ y $b>0$, la función de distribución de X toma la forma

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

203. Sea X una v.a. con distribución beta(a,b). Denuestre que

$$E(X^n) = \frac{B(a+n,b)}{B(a,b)}.$$

204. Sea X una v.a. con distribución beta(a,b), con $a=b=1$. Demuestre que X tiene una distribución uniforme en el intervalo $(0,1)$.

205. Sea X una v.a. con distribución $\mathrm{N}(\mu,\sigma^2)$. Verifique que $f_X(x)$ es efectivamente una función de densidad.

206. Sea X una v.a. con distribución $\mathrm{N}(\mu,\sigma^2)$. Demuestre que

a) $E(X)=\mu$.

b) $\mathrm{Var}(X)=\sigma^2$.

207. Sea X con distribución $\mathrm{N}(\mu,\sigma^2)$. Demuestre que la variable

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

tiene una distribución normal estándar.

208. Sea Z una v.a. con distribución normal estándar. Demuestre que la v.a. $X=\mu+\sigma Z$ tiene una distribución $\mathrm{N}(\mu,\sigma^2)$.

209. Sea X una v.a. 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) $P(-20 < X \leq 10)$.

210. Sea X una v.a. con distribución $\mathrm{N}(0, 100)$. Calcule

   a) $P(X \leq 10)$.
   b) $P(X < 0)$.
   c) $P(0 < X \leq 100)$.
   d) $P(X \geq 30)$.
   e) $P(-10 < X \leq 10)$.

211. Encuentre x tal que $\Phi(x) = 0.8666$.

212. Encuentre x tal que $1 - \Phi(x) = 0.9154$.

213. 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.

214. 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.

215. Sea X una variable aleatoria con distribución $\mathrm{N}(0, 1)$. Demuestre que X^2 tiene una distribución $\chi^2(1)$.

216. Sea X una variable aleatoria con distribución normal estándar. Encuentre la función de densidad de $Y = |X|$.

217. 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?

218. Enuncie con precisión el teorema central del límite.

219. Sea X una v.a. con distribución $\chi^2(n)$. Verifique que $f_X(x)$ es efectivamente una función de densidad.

220. Sea X una v.a. con distribución $\chi^2(n)$. Demuestre que
   a) $E(X) = n$.
   b) $\mathrm{Var}(X) = 2n$.

221. Sea X una v.a. con distribución t. Verifique que $f_X(x)$ es efectivamente una función de densidad.

222. Sea X una v.a. con distribución t. Demuestre que
   a) $E(X) = 0$.
   b) $\mathrm{Var}(X) = \frac{n}{n-2}$ para $n > 2$.

Vectores aleatorios

223. 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?

e) Calcule $E(XY)$.

224. 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?

e) Calcule $E(XY)$.

225. 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?

e) Calcule $E(XY)$.

226. 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?

e) Calcule $E(XY)$.

Estadística descriptiva

227. Calcule la media, moda, mediana, varianza y desviación estándar del siguiente conjunto de datos: $4, 2, 0, 9, 4, 2, -1, 1, -4, 2$.

228. Pregunte a diez personas sus estaturas. Registre todos estos datos y calcule la media, moda, mediana, varianza y desviación estándar.

229. 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 y desviación estándar.

230. Escriba sus ultimas diez calificaciones semestrales y calcule la media, moda, mediana, varianza y desviación estándar.

231. Demuestre que $s^2 = \dfrac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \dfrac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$.

232. Demuestre que $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.

233. Demuestre que $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$.

234. Encuentre el valor de c que minimiza la función $\bar{x}(c) = \sum_{i=1}^{n}(x_i - c)^2$.

235. Calcule la media, varianza, deviación 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

236. Calcule la media, varianza, desviación estándar, mediana y moda aproximados del siguiente conjunto de datos agrupados. Elabore además un histograma.

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

Estimación puntual

237. ¿Qué es un estimador puntual?

238. ¿Cuándo decimos que un estimador es insesgado?

239. Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados para el parámetro θ. Demuestre que $\hat{\theta} = \alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2$ es un nuevo estimador insesgado para θ, en donde α es una constante.

240. Sea $X_1, X_2, \ldots, X_n$ una m.a. de una población con media conocida μ y varianza finita σ^2 desconocida. Demuestre que

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

es un estimador insesgado para σ^2.

241. Sea $X_1, X_2, \ldots, X_n$ una m.a. de una población con media desconocida y varianza finita σ^2 desconocida. Demuestre que

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

es un estimador insesgado para σ^2.

242. Sea $X_1, X_2, \ldots, X_n$ una m.a. de una población con media desconocida y varianza finita σ^2 desconocida. Demuestre que

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2$$

es un estimador insesgado para σ^2.

243. Sea $X_1, X_2, \ldots, X_n$ una m.a. de una población con media μ y varianza σ^2 desconocidas. Recordemos que la varianza muestral S^2 se ha definido como sigue $S^2 = \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n} (X_i - \bar{X})^2$. Demuestre que S no es un estimador insesgado para σ. Modifique S para que sea insesgado. [Sugerencia: $\dfrac{(n-1)}{\sigma^2} S^2 \sim \chi^2(n-1)$].

244. Dada una muestra aleatoria de tamaño n de una población uniforme en el intervalo (a, b) en donde $a = 0$, use el método de momentos para encontrar un estimador para el parámetro b.

245. 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.

246. 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.

247. Dada una muestra aleatoria de tamaño n de una población normal con parámetros desconocidos μ y σ^2, use el método de momentos para encontrar estimadores para los parámetros μ y σ^2 .

248. Dada una muestra aleatoria de tamaño n de una población uniforme en el intervalo (a, b) en donde $a = 0$, use el método de máxima verosimilitud para encontrar un estimador para el parámetro b.

249. Dada una muestra aleatoria de tamaño n de una población Poisson con parámetro desconocido $\lambda > 0$, use el método de máxima verosimilitud para encontrar un estimador del parámetro λ.

250. Dada una muestra aleatoria de tamaño n de una población exponencial con parámetro desconocido $\lambda > 0$, use el método de máxima verosimilitud para encontrar un estimador del parámetro λ.

251. Dada una muestra aleatoria de tamaño n de una población normal con parámetros desconocidos μ y σ^2, use el método de máxima verosimilitud para encontrar estimadores para los parámetros μ y σ^2 .

Estimación por intervalos

252. ¿Qué es un intervalo de confianza?

253. 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.

254. Se realizan 20 pruebas de resistencia de un cierto material obteniendose 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

255. ¿Qué es una hipótesis estadística?

256. ¿Qué es una hipótesis simple?

257. ¿Qué es una hipótesis compuesta?

258. ¿Qué es una prueba de hipótesis?

259. ¿Qué es el error tipo I en una prueba de hipótesis?

260. ¿Qué es el error tipo II en una prueba de hipótesis?

261. ¿Qué es la región crítica en una prueba de hipótesis?

262. ¿Qué es el nivel de significancia en una prueba de hipótesis?

263. ¿Qué se entiende por tamaño de la región crítica en una prueba de hipótesis?

264. 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.

265. Se cree que la estatura promedio de los mexicanos es de 1.70 metros de estatura. Lleve a cabo 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, 1.66, 1.83, 1.77, 1.74, 1.64, 1.69, 1.72, 1.66, 1.55, 1.60, 1.62.

266. Considere la prueba de hipótesis para la media μ de una población normal con σ^2 conocida. Demuestre que la probabilidad de cometer el error tipo II, esto es, $\beta(\mu_1)$ tiende a cero cuando el tamaño de la muestra n tiende a infinito.

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] Wisniewski P. M., Bali G. *Ejercicios y problemas de la teoría de las probabilidades.* Trillas, 1998.

[15] Velasco G., Wisniewski P. *Probabilidad y estadística para ingeniería y ciencias.* Thomson, 2001.

Índice