Cómo medir la incertidumbre con probabilidades

CAPÍTULO 6

6.1. INTRODUCCIÓN

Tomamos muestras de poblaciones, así que nuestras conclusiones o inferencias sobre la población tendrán cierta cantidad de incertidumbre. Ya nos hemos familiarizado con la idea de probabilidad, hablamos de la chance de un error tipo I en un proceso de decisión.

Sabemos que el valor-p es una medida de la posibilidad de ocurrencia del valor observado calculada bajo la hipótesis nula. Vimos que la aleatorización juega un papel importante en el muestreo y en la asignación de unidades en los experimentos. Aprendimos que un modelo provee un resumen muy útil de la distribución de una característica que sirve de marco de referencia para la toma de decisiones bajo condiciones de **incertidumbre**.

Los juicios probabilísticos son parte de nuestra vida diaria. Sin duda habrán escuchado comentarios como los siguientes:

- *No tengo chance de aprobar el "quiz" de la clase de mañana.*
- *El juez de línea revoleó una moneda para saber quién comienza el juego, cada jugador tiene chance de 50 contra 50 de sacar primero.*

¿Qué quiere decir tener una chance de 50 contra 50?

Empezamos nuestra aventura con la probabilidad, estudiando cómo obtener probabilidades mediante la simulación. Después, estudiaremos los conceptos básicos de la probabilidad y las reglas que debemos aplicar para obtener probabilidades de varios eventos. Finalmente reveremos las ideas del capitulo "Modelizar para decidir" usando un modelo para la distribución de una variable, donde ahora, a la variable la llamaremos **variable aleatoria** y al modelo asociado **distribución de probabilidad**.

6.2 . ¿QUÉ ES LA PROBABILIDAD?

Tenemos una moneda, de un lado es "cara" y del otro "cruz". Una moneda se dice correcta si la chance de obtener cara es igual a la chance de obtener cruz. Vamos a arrojar la moneda, ¿por qué diremos que la probabilidad de cara es 1/2? ¿Qué quiere decir? Si arrojamos una moneda correcta un gran número de veces, esperamos que la mitad de las veces resulte cara. Este uso de la palabra probabilidad basada en la **interpretación de la frecuencia relativa**, se aplica a situaciones en las que las condiciones son exactamente repetibles. La **probabilidad** de un resultado se define como la proporción de veces que ocurriría el evento si el proceso se repitiera muchas veces bajo las mismas condiciones. Esto se llama también la **frecuencia relativa del largo plazo** del resultado.

Definición
La **probabilidad** de que ocurra un suceso es la proporción de veces que ocurre en el largo plazo; es decir, la frecuencia relativa con que ocurre el evento.

El énfasis en el **largo plazo** o **muchas repeticiones** es muy importante. La probabilidad de que una cara es igual a 1/2 no significa que se obtendrá una cara cada dos tiradas de la moneda. Arrojar la moneda cuatro veces y observar la secuencia XCXX no es evidencia de que la probabilidad de cara es 1/4. Sin embargo, si en 1.000 tiradas, aproximadamente el 25% de los resultados es cara, sería razonable concluir que la moneda es viciada y que la probabilidad de cara es 0.25. A medida que aumenta el número de repeticiones, esperamos que la proporción de caras se empiece a estabilizar alrededor de un valor constante. Este es el valor que tomamos como probabilidad de cara.

El enfoque de las frecuencias relativas para definir probabilidades, se aplica en situaciones que se pueden pensar como repetibles bajo condiciones semejantes. Algunas situaciones no son de este tipo.

Si estamos planeando una fiesta al aire libre para el próximo sábado a la tarde, ¿cuál es la probabilidad de que llueva?

Dos equipos de fútbol, Central y Newell´s, juegan la final de un torneo ¿cuál es la probabilidad de que gane Central?

En estas situaciones, una persona puede usar sus propias experiencias y conocimientos para asignar una probabilidad al suceso. Tales

probabilidades se llaman probabilidades personales o subjetivas, representan el grado de creencia personal en la ocurrencia del suceso. Personas distintas pueden asignar probabilidades distintas, y todas pueden ser correctas. Cualquier probabilidad, sin embargo, debe estar entre 0 y 1 (o entre 0% y 100%).

Las probabilidades nos ayudan a tomar decisiones. Si la noche del viernes anterior a la fiesta, el pronóstico del tiempo para el sábado es que se sucederán períodos de lluvia y de mucho calor, uno puede decidir hacer la fiesta bajo techo. Necesitamos viajar a Buenos Aires por cuestiones de negocio en un vuelo que parte el martes a la mañana para llegar a media mañana. Hay dos aerolíneas que ofrecen vuelos en ese horario, una llega a horario el 88% de las veces, la otra el 73%. Estas probabilidades, junto con otros datos como precio del pasaje, ayudan a decidir en cuál de las dos aerolíneas reservar asiento. Sin embargo, sin importar cuál es la elegida, ese vuelo particular puede llegar a horario o no. Las probabilidades no pueden determinar si un suceso va a ocurrir o no en un caso individual.

6.3. SIMULANDO PROBABILIDADES

Uno de los conceptos básicos del estudio de las probabilidades es un **proceso aleatorio** o **fenómeno aleatorio**. Un proceso es aleatorio cuando se lo puede repetir bajo condiciones similares. Si bien se conoce el conjunto de resultados posibles, en una repetición individual no se puede predecir con seguridad cuál será el resultado exacto. Sin embargo, hay un patrón de comportamiento predecible en el largo plazo, tal como que la frecuencia relativa de un resultado dado se acerque a un valor constante. Arrojar una moneda es un ejemplo de un proceso aleatorio.

Definición
Un proceso o experimento aleatorio es un proceso repetible del que se conoce el conjunto de resultados posibles, pero no puede predecirse con seguridad un resultado exacto. Sin embargo, hay un patrón de comportamiento predecible en el largo plazo, tal como que la frecuencia relativa de un resultado dado se acerque a un valor constante.

Arrojar una moneda al aire 10 veces dio como resultado la secuencia CXXCCCCXXC. Esta secuencia tiene cuatro caras consecutivas (una corrida de cuatro caras). ¿Se puede considerar inusual una corrida de cuatro caras si la moneda es correcta? ¿Cuál es la probabilidad de obtener una corrida de cuatro caras si arrojamos diez veces una moneda correcta? Encontrar una probabilidad, ya sea repitiendo el proceso aleatorio o usando las matemáticas más complejas del cálculo de probabilidades, puede ser difícil.

De modo alternativo, se puede estimar la probabilidad de un evento a través de la **simulación**. Simular significa imitar, generar condiciones que se aproximen a las actuales. Para simular un proceso aleatorio se puede usar diferentes artificios: calculadora, computadora, o una tabla de números aleatorios. Primero se necesita especificar las condiciones del fenómeno aleatorio (proveer un modelo que liste todos los resultados posibles y sus respectivas probabilidades). Para el caso de arrojar una moneda correcta, los sucesos elementales son cara y cruz, y la probabilidad de cada uno es 1/2. A continuación necesitamos detallar cómo simular un suceso elemental (es decir, establecer cuál resultado de la simulación corresponde a cada suceso elemental).

Nuestro problema original se refiere a arrojar 10 veces una moneda correcta. Debemos simular diez tiradas para representar una única repetición del proceso. Finalmente, podríamos simular una gran cantidad de veces y determinar el número de veces que ocurrió el evento de interés. La frecuencia relativa, el número de veces que el evento ocurre sobre la cantidad total de simulaciones, se usa para estimar la probabilidad del evento.

Definición
Una simulación es la imitación del comportamiento del azar usando artificios como generadores de números aleatorios o tablas de números al azar.

Los **pasos básicos** para calcular una probabilidad por simulación son:

- Especificar un modelo para los sucesos elementales y el fenómeno aleatorio subyacente.
- Delinear cómo simular un suceso elemental y cómo representar una repetición del proceso aleatorio.
- Simular muchas repeticiones y estimar la probabilidad de un evento con su frecuencia relativa.

☑ **PARA RESOLVER !!!**

6.1. Un Plan Familiar

Una pareja planea tener hijos. Ellos quisieran tener un varón para que lleve el apellido de la familia. Después de una discusión, deciden tener tres hijos hasta que llegue un varón o hasta tener tres mujeres, lo que ocurra primero. ¿Cuál es la probabilidad de que tengan un varón? Vamos a simular el plan familiar de esta pareja y estimaremos esa probabilidad.

Paso 1: Especificar un modelo para los resultados individuales

El proceso aleatorio es tener hijos hasta que nazca un varón o nazcan tres hijas, lo primero que ocurra. El fenómeno aleatorio individual es "tener un hijo" y la respuesta de interés es su género. Necesitamos empezar con algunos supuestos básicos del resultado de este experimento aleatorio. Parece razonablemente justo suponer que:

- cada hijo tiene un 50% de probabilidad de ser varón o mujer,
- el género de los sucesivos hijos es independiente, es decir, conocer el género de uno de los hijos no influye en el género del siguiente.

Paso 2: Simular resultados individuales y una repetición

Necesitamos simular el género de un hijo. Podemos usar la calculadora o una computadora con un generador de números aleatorios. Hay sólo dos resultados posibles, varón o mujer, así que necesitamos generar una secuencia aleatoria de dos valores (por ejemplo 1 y 2). Necesitamos determinar cuál valor representa varón y cuál mujer:

Sea 1 = nació un varón, entonces . . . . = nació una mujer.

Para simular una repetición del plan familiar, generamos números aleatorios hasta que obtengamos un varón o tres mujeres.

Si usamos una tabla de números aleatorios, tenemos 10 dígitos, del 0 al 9. Necesitamos decidir 5 números que simulen, por ejemplo, el nacimiento de un varón:

Sea {0, 1, 2, 3, 4} = {nació un varón}, entonces . . . . = {nació una mujer}

Comenzamos en la línea 14 de la Tabla I, leyendo de izquierda a derecha, los primeros dígitos se reproducen debajo, una "V" o una "M" indica varón o mujer respectivamente, una línea separa las repeticiones sucesivas.

9	8	3	|	6	1	|	5	9	2	|
M	M	V	|	M	V	|	M	M	V	|

fin de la primera repetición

(a) En la primera repetición ¿cuántos hijos tuvo la pareja? TRES
¿Tuvieron el varón? SÍ
(b) En la segunda repetición ¿cuántos hijos tuvo la pareja? . .
¿Tuvieron el varón? . . . .
(c) En la tercera repetición ¿cuántos hijos tuvo la pareja? . . . .
¿Tuvieron el varón? . . . .

También se puede usar un dado no cargado y asignar tres resultados, por ejemplo 1, 3, 5 para representar el evento "varón".

Paso 3: Simular muchas repeticiones y estimar la probabilidad

Trabajando con un compañero, o en grupos pequeños, simule muchas repeticiones (por lo menos 10) del plan familiar, y use la frecuencia relativa del evento "la pareja tiene un varón" para estimar la probabilidad. Comience escribiendo una lista de números aleatorios, debajo de cada valor escriba "V" o "M" para representar un varón o una mujer, agregue una línea para separar las repeticiones sucesivas. Debe generar números suficientes como para tener 10 líneas que correspondan a 10 repeticiones.

(a) De acuerdo con sus 10 repeticiones, ¿cuántas veces la pareja tuvo un varón? . . . .
(b) Su estimación de la probabilidad de que esta estrategia produzca un varón es . . . .

La estimación que hizo en el paso anterior usando la frecuencia relativa no es muy precisa porque sólo llevo a cabo 10 repeticiones. Combine las frecuencias de todos los grupos de la clase y dé una estimación de la probabilidad de que la estrategia produzca un varón.

Grupo	# repeticiones	# veces que nace un varón
1		
2		
3		
4		
5		
6		
7		
8		
TOTAL	N =	#B =

Así es que nuestra estimación combinada de la probabilidad de que esta estrategia produzca un varón es:

$$= \frac{\# B}{N} = \dots\dots\dots\dots\dots\dots$$

Más adelante aprenderemos cómo calcular matemáticamente la verdadera probabilidad de tener un varón, que es 0.875.

☑ **PARA RESOLVER !!!**

6.2. Simulando otros eventos

(a) ¿Cómo simular un evento que tiene una probabilidad de ocurrencia 0.4?
Sea . . . . . = ocurre el evento; . . . . . . = el evento no ocurre.
(b) ¿Cómo se puede simular un evento con cuatro resultados posibles: A, B, C y D, con probabilidad 0.1, 0.2, 0.3 y 0.4 respectivamente?
Sea:
. . . . = ocurre A . . . . = ocurre B
. . . . = ocurre C . . . . = ocurre D

(c) ¿Cómo simular un evento que tiene probabilidad de ocurrencia 0.45?
Sea: . . . . = ocurre el evento; . . . . . = el evento no ocurre

☑ **PARA RESOLVER !!!**
6.3. ¿Un máximo de 13?

(a) Suponga que en un país muy particular a las parejas se les permite tener un solo hijo. Nos referiremos a esta situación como la regla original.

Bajo esta regla original,
- ¿Cómo será el número de varones comparado con el de mujeres?
- ¿Cuál será el número promedio de hijos por familia?

(b) Considere una regla nueva
"Las parejas pueden tener hijos hasta que tengan un varón o hasta que tengan 13 hijos"

¿Cómo influiría esta nueva regla en la cantidad de varones y mujeres del país?
¿Cuál sería el número promedio de hijos por familia?

(c) Lleve a cabo una simulación y presente los resultados.

☑ **PARA RESOLVER !!!**
6.4. Las tres puertas

Hay tres puertas. Detrás de una hay un auto. Detrás de las otras dos hay una cabra. Una vez que haya elegido una puerta usted recibe el premio que está detrás de la puerta.

El conductor del juego sabe qué hay detrás de cada puerta. Cuando uno elige una puerta, el conductor abre una de las dos restantes; siempre hay una con una cabra detrás para abrir.

Se tienen dos opciones:
1) Mantener la puerta que se eligió originalmente.
2) Cambiar por la puerta cerrada restante.

¿Qué elegiría?

Hagamos una simulación para estimar la probabilidad de ganar si conserva la primera puerta elegida y la probabilidad de ganar si cambia.

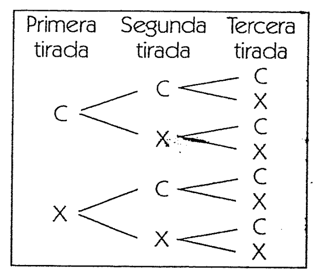
6.4. EL LENGUAJE DE LAS PROBABILIDADES

En esta sección retomamos algunas ideas básicas sobre probabilidad e introduciremos algunas reglas y notación. Las reglas nos permitirán calcular las probabilidades de eventos simples y compuestos. Comenzamos por enumerar las principales componentes de un estudio formal de las probabilidades.

6.4.1 Espacios Muestrales

Primero, tenemos un **experimento aleatorio**. Puede ser arrojar tres veces una moneda, tirar un par de dados o elegir al azar un futuro votante en una encuesta. A continuación tenemos el **espacio muestral** o **conjunto de resultados** del experimento aleatorio. El espacio muestral, llamado con S, es el conjunto de resultados posibles de un experimento aleatorio.

Si el experimento aleatorio consiste en arrojar tres veces una moneda correcta, entonces los resultados que conforman el espacio muestral se pueden obtener en forma ordenada usando el método del árbol como se muestra a continuación.



S = CCC, CCX, CXC, CXX, XCC, XCX, XXC, XXX

ó **S** = {CCC, CCX, CXC, CXX, XCC, XCX, XXC, XXX}

Nota: Para separar cada resultado de una lista se usa una coma

Hay ocho sucesos individuales posibles en este espacio muestral. Como suponemos que la moneda es correcta, los ocho son igualmente probables (es decir que la probabilidad de cada suceso elemental es 1/8).

Si el experimento aleatorio es arrojar 3 veces una moneda y se define el suceso como el número de caras, entonces el espacio muestral está dado por:

$$S = \{0,1,2,3\}$$

Hay cuatro resultados posibles en este último espacio muestral. Sin embargo, estos cuatro resultados no son igualmente probables. Obtener exactamente una cara es más probable que no obtener ninguna, ya que al resultado exactamente una cara corresponde tres sucesos elementales {CXX, XCX, XXC} y sólo uno {XXX} corresponde a ninguna cara.

A partir de estos dos ejemplos podemos ver que:

- El espacio muestral no tiene que ser necesariamente un conjunto de números, sin embargo, se puede establecer un código si los resultados no son numéricos;
- La definición de **qué** es un resultado individual es fundamental para representar el espacio muestral en forma correcta;
- Los sucesos elementales de un espacio muestral no son necesariamente equiprobables.

Definición
Un **espacio muestral** o **conjunto de resultados posibles** es el conjunto de todos los sucesos elementales de un proceso aleatorio. Comúnmente se lo denomina con **S** y se lo puede representar con una lista, un árbol, un intervalo de valores.

☑ **PARA RESOLVER !!!**
6.5. Espacios muestrales

Escriba el espacio muestral de las siguientes experiencias. Se dan algunas como ejemplo.

(a) Tirar una moneda de dos caras una sola vez: S = {C , X }

(b) Tirar dos veces un dado no cargado:

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

(c) Tirar dos dados y registrar la suma de los valores de los dos dados.

S = {

(d) Tomar una muestra aleatoria de tamaño 10 de un lote de piezas y registrar el número de defectuosos en la muestra.

S = {

(e) Seleccionar un estudiante al azar y registrar el tiempo que dedicó a estudiar Estadística en las últimas 24 horas.

S = [0 , 24]

(f) El tiempo mientras que una persona llega a la parada del colectivo y la llegada del colectivo.

S = [

☑ **PARA RESOLVER !!!**
6.6. Las preferencias de los votantes

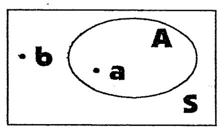
Considere el proceso de elegir aleatoriamente dos adultos de la ciudad de Rosario y preguntarle a cada uno si es justicialista, radical, socialista o de algún otro partido político. Los dos adultos seleccionados son Juan y Pablo. ¿Cuál de los siguientes es el espacio muestral correcto para la serie de resultados posibles de este experimento?

a) S = Juan , Pablo
b) S = Justicialista, Radical, Socialista, Otro
c) S = Justicialista , Socialista
d) Ninguno de los anteriores.

¿Seleccionó la respuesta **(b)** en el ejercicio precedente? Si es así, eligió el espacio muestral correcto para el proceso aleatorio "elegir aleatoriamente una persona de Rosario y registrar su preferencia política".

¿Seleccionó la respuesta **(c)**? Si es así, eligió un conjunto que representa sólo uno de los posibles resultados individuales. La respuesta correcta es la **(d)**, ya que el espacio muestral del proceso definido contiene 16 sucesos elementales. Destaquemos de paso que el resultado (J, R) "Juan es justicialista y Pablo es radical" es diferente del (R, J) "Juan es radical y Pablo es justicialista". En otras palabras el orden de las respuestas interesa. Si estuviéramos investigando una gran cantidad de adultos para conocer su proporción en cada categoría de preferencia política, el orden de las respuestas no interesaría.

Un subconjunto del espacio muestral se llama **evento** y se lo denota con las letras mayúsculas del comienzo del alfabeto (A, B, C, etc.). En algunos casos el espacio muestral y los eventos se representan usando diagramas de Venn. El espacio muestral se representa con un rectángulo y los eventos como un subconjunto del rectángulo.



Suponga que se lleva a cabo el experimento y que el resultado es "a". Como el resultado "a" está en el evento A, se dice que ocurrió el evento A. Si el resultado es "b", como "b" no está en A, decimos que no ocurrió el evento A.

Si el experimento aleatorio fuera tirar un dado correcto, el espacio muestral sería

S = {1,2,3,4,5,6}

Definimos el evento A como "resultado impar". Entonces el evento **A** = {1,3,5} es un subconjunto de S. Si se tira el dado y sale "1", entonces ocurrió el evento A. Si se tira el dado y sale "6" entonces el evento A no ocurrió.

Definición
Un **evento** es cualquier subconjunto del espacio muestral S. Se dice que ha ocurrido un evento A si, al repetirse una vez el proceso aleatorio, ocurre cualquiera de los resultados de A.

☑ **PARA RESOLVER !!!**

6.7. Definir Eventos

La experiencia consiste en tirar dos dados.
Encierre con un círculo los resultados correspondientes a los siguientes eventos:

(a) A = "No se obtiene ningún seis"

S = { (1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6) }

A=

(b) B = "Exactamente un seis"

S = { (1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6) }

(c) C = "Exactamente dos seis"

S = { (1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6) }

(d) D = "Al menos un seis"

S = { (1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6) }

☑ PARA RESOLVER !!!

6.8. A favor o en contra

En un grupo de gente, algunos están a favor (F) del aborto y otros en contra (C). Se seleccionan tres personas al azar de este grupo y se anotan sus opiniones a favor o en contra del aborto.

(a) Escriba el espacio muestral correspondiente
S =

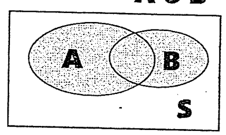
(b) Escriba los resultados que conforman el evento A = "al menos una persona está en contra"
A =

(c) Escriba los resultados que conforman el evento B = "exactamente dos personas están a favor"
B =

A veces estamos interesados en eventos que no son tan simples. Los eventos pueden ser combinaciones de varios eventos. A continuación veremos algunas de estas posibles combinaciones y la relación entre los eventos.

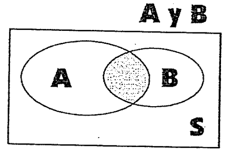
Unión de dos eventos: representada por **A o B**, matemáticamente se escribe **A∪B**, corresponde la zona sombreada del diagrama siguiente. La unión A o B contiene los resultados que están en el evento A o en el B o en ambos. A veces al evento A o B se lo identifica con *"al menos uno de los dos"* eventos ha ocurrido.

A o B



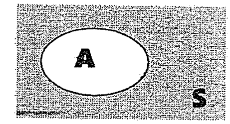
Intersección de dos eventos: representada por **A y B**, se escribe matemáticamente **A∩B**, corresponde la zona sombreada del siguiente diagrama. La intersección esta formada sólo por aquellos resultados que están en ambos eventos, el A y el B. A menudo se usa la palabra *"ambos"* para describir la intersección de dos eventos.

A y B

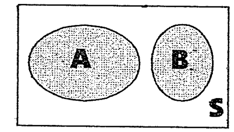


Complemento de un evento: representado por **no A**, se escribe matemáticamente $\mathbf{A^c}$, corresponde la zona sombreada del siguiente diagrama. El complemento de un evento A está formado por todos los resultados que no están en el evento A. Listar los elementos que componen un evento A parece a veces algo excesivo, en esos casos es mejor listar los resultados que forman parte del complemento. Todo evento A tiene un único evento complementario A^c en S.

$\mathbf{A^c}$



Dos eventos se dicen **disjuntos** si no tienen resultados en común. Algunas veces se los llama **mutuamente excluyentes**. En notación matemática escribiríamos **A∩B = ∅**, donde ∩ representa intersección y ∅ representa conjunto vacío, el conjunto que no contiene resultados. Dos eventos son disjuntos si no pueden ocurrir al mismo tiempo. Los eventos disjuntos se pueden representar con diagramas de Venn. El diagrama de la derecha muestra dos eventos A y B que son disjuntos.



Definición
Dos eventos A y B son **disjuntos** o **mutuamente excluyentes** si no tienen resultados en común. Así, si uno de ellos ocurre, el otro no puede ocurrir.

La noción de eventos mutuamente excluyentes se puede extender a más de dos eventos. Por ejemplo diremos que los eventos A, B y C son mutuamente excluyentes si:

1. A, B y C no tienen resultados en común, es decir, $A \cap B \cap C = \emptyset$
2. los eventos A y B no tienen resultados en común, $A \cap B = \emptyset$
   los eventos A y C no tienen resultados en común, $A \cap C = \emptyset$
   los eventos B y C no tienen resultados en común, $B \cap C = \emptyset$

EJEMPLO 1. Eventos disjuntos

	Primario	Secundario	Terciario
Varón	38	28	22
Mujer	45	50	17

Una muestra de 200 adultos se clasificó por género (88 varones y 112 mujeres) y máximo nivel educativo alcanzado (83 primario, 78 secundario y 39 terciario).

Considere los eventos:

A = "el adulto elegido tiene nivel terciario"

B = "el adulto elegido es varón y el máximo nivel educativo alcanzado es secundario"

C = "el adulto elegido es mujer"

Como ningún adulto puede tener a la vez como máximo nivel educativo el secundario y el terciario, A y B son disjuntos. Sin embargo, un adulto puede a la vez ser mujer y tener como máximo nivel educativo el secundario, los eventos A y C no son disjuntos. Las dos categorías de las filas (Varón y Mujer) son los niveles de la variable Género, y son disjuntas. Las tres categorías de las columnas (Primario, Secundario y Terciario) son los niveles para el más alto nivel educativo alcanzado, y son disjuntas.

☑ **PARA RESOLVER !!!**

6.9 ¿Mutuamente excluyentes?

Para cada escenario, determine si cada una de las siguientes listas de eventos es mutuamente excluyente:

a) Se selecciona un adulto al azar de una población.
A = "el adulto seleccionado es varón"
B = "el adulto seleccionado tiene más de 55 años de edad"
No, hay un resultado posible en común, el adulto seleccionado "es un varón de más de 55 años".

b) Se hace una venta en un comercio minorista.
A = "el monto de la venta excede $50"
B = "el monto de la venta excede $500"

c) Se hace una venta en un comercio minorista.
A = "el monto de la venta es menor de $50"
B = "el monto de la venta está entre $100 y $450"
C = "el monto de la venta excede $1.000"

d) Se eligen diez estudiantes al azar.

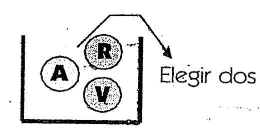
A = "las mujeres no son más de 3"
B = "hay al menos 7 mujeres"
C = "como máximo 5 son mujeres"

EJERCICIOS

6.1. Se seleccionará una muestra aleatoria de estudiantes de la facultad y se registrará si cada uno de ellos trabaja o no. Escriba el espacio muestral correspondiente a:

a) seleccionar un solo estudiante,.
b) seleccionar cuatro estudiantes.
c) seleccionar 20 estudiantes, y el resultado de interés es el número de estudiantes que no trabajan.

6.2. Reemplazo y orden: Una canasta contiene tres pelotas, una verde, una amarilla y otra blanca. Se seleccionan dos. Por ejemplo, el resultado "V, A" representa que primero se eligió la pelota verde, seguida por la amarilla.



Escriba el espacio muestral correspondiente, si...

a) se muestreó **con** reemplazo y el **orden interesa**.
b) se muestreó **con** reemplazo y el **orden no interesa**.
c) se muestreó **sin** reemplazo y el **orden interesa**.
d) se muestreó **sin** reemplazo y el **orden no interesa**.

6.3. ¿Disjuntos? Considere le experiencia de sacar una carta de un mazo de baraja española.
Sea A = "Basto" , B = "Rey" y C = "Espada".

¿Son disjuntos los eventos A y B? Explique.
a) ¿Son disjuntos los eventos A y C? Explique.
b) ¿Son disjuntos los eventos B y C? Explique.

6.4. Un minorista de software ha hecho un inventario del soft del año anterior para ponerlo en liquidación. De los 380 paquetes, 270 se pueden usarse desde el CD-ROM y el resto debe instalarse en el rígido. De los de CD-ROM, 200 son compatibles con Mackintosh, 180 se usan en sistemas de PC, y 110 son compatibles tanto en PC como en Mackintosh.

a) Use un diagrama de Venn para representar la información. Asegúrese de dar nombre a los eventos.

b) Los eventos "CD-ROM" y "Disco rígido" ¿son disjuntos?

c) Los eventos "Mackintosh" y "Sistemas PC" ¿son disjuntos?

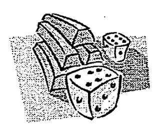
6.4.2 Reglas de Probabilidad

Retomamos la idea de probabilidad y su relación con los resultados de un espacio muestral. A cualquier evento A se le asigna un número P(A) llamado **probabilidad del suceso A**. Recuerde que la probabilidad de un evento se define como la frecuencia relativa con que ese evento puede ocurrir. Cuando el espacio muestral contiene un número finito de resultados posibles hay otra técnica para asignar probabilidades a los eventos:

- asignar una probabilidad a cada uno de los resultados individuales o elementales, es decir, un número entre 0 y 1, tal que su suma dé 1, y
- la probabilidad de cualquier evento es la suma de las probabilidades de los resultados que componen dicho evento.

Si los sucesos del espacio muestral son equiprobables la probabilidad de un evento A es simplemente la proporción de resultados del espacio muestral que forman el evento A. No suponga automáticamente que todos los resultados tienen la misma probabilidad de ocurrir, esto depende del proceso aleatorio y de la definición del suceso.

☑ **PARA RESOLVER !!!**



6.10. Asignar probabilidades a los eventos

Considere el experimento de arrojar dos dados sin fallas. En el espacio muestral hay 36 puntos igualmente probables. ¿Cuáles son las probabilidades de los siguientes eventos?

a) Evento A = "ningún seis"

S = { (1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)

(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)}

Como 25 de los 36 resultados igualmente probables componen el evento A, tenemos:

$$P(A) = \frac{25}{36}$$

b) Evento B = " exactamente un seis" P (B) =

c) (c) Evento C = " exactamente dos seis"   P (C) =

d) (d) Evento D = " al menos un seis" P (D) =

e) Compare 1 - P (A) con P (D). Los eventos A y D son complementarios.

f) Considere la suma de los valores de dos dados. ¿Cuál es la probabilidad de obtener una suma de tres en la próxima tirada? ¿Cuál es la probabilidad de obtener una suma de al menos 11 en la próxima tirada?

REGLAS BÁSICAS
para asignar probabilidades a eventos

6.1. $0 \le P(A) \le 1$ Una probabilidad es siempre un número entre cero y uno. La probabilidad es cero si el suceso no puede ocurrir (suceso imposible). La probabilidad es uno si el suceso ocurre con seguridad.

6.2. $P(S) = 1$ Si sumamos las probabilidades de cada uno de los resultados individuales del espacio muestral, la probabilidad total es uno.

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

La tercera regla se llama **regla del complemento**. Cualquier evento y su respectivo complemento son conjuntos disjuntos, si los unimos tendremos todo el espacio muestral S. La probabilidad del espacio muestral S es 1, así que las probabilidades del evento y su complemento suman 1. Esta regla es especialmente **útil**. Si hallar la probabilidad de un suceso parece muy difícil, conviene probar si el cálculo de la probabilidad del complemento es más fácil.

PARA PENSAR !!!

Considere el experimento de tirar una moneda 10 veces. Sea A el evento "al menos una cara". Al menos una cara es 1 cara o 2 caras o 3 caras o 4 caras o 5 caras o 6 caras o 7 caras o 8 caras o 9 caras o 10 caras. Son un montón de eventos para tratar de contarlos.

Piense en cuál es el complemento de A y luego encuentre la probabilidad del evento A usando la regla del complemento.

☑ **PARA RESOLVER !!!**

6.11. ¿Dado correcto?

Se sospecha que un dado está cargado con tendencia a mostrar los valores más altos. Deseamos testar las siguientes hipótesis:

H_0: El dado es correcto (todas las caras tienen la misma chance)

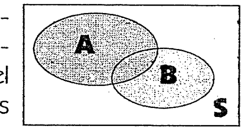
H_1: El dado tiene tendencia a mostrar los valores mayores.

Tiraremos el dado dos veces. Los 36 resultados posibles son:

$$
\begin{aligned}
S = \{ & (1,1)\ (1,2)\ (1,3)\ (1,4)\ (1,5)\ (1,6) \\
& (2,1)\ (2,2)\ (2,3)\ (2,4)\ (2,5)\ (2,6) \\
& (3,1)\ (3,2)\ (3,3)\ (3,4)\ (3,5)\ (3,6) \\
& (4,1)\ (4,2)\ (4,3)\ (4,4)\ (4,5)\ (4,6) \\
& (5,1)\ (5,2)\ (5,3)\ (5,4)\ (5,5)\ (5,6) \\
& (6,1)\ (6,2)\ (6,3)\ (6,4)\ (6,5)\ (6,6)\}
\end{aligned}
$$

La regla de decisión sugerida es: **Rechazar H_0** si la suma de los dos dados es demasiado grande. ¿Cuál es el valor de p asociado a observar un resultado de 11?

A o B



Nuestra siguiente regla básica nos dice cómo encontrar la probabilidad de la unión de dos eventos; es decir, la probabilidad de que ocurra un evento o el otro. Es fácil comprender esta regla mirando el diagrama correspondiente. Comenzamos tomando los resultados del evento A y le agregamos todos aquellos resultados que conforman el evento B. Como los resultados que ocurren tanto en A como en B los hemos incluido dos veces, los restamos una.

La Regla de la Suma

6.4. P (ocurra al menos uno, A ó B) = **P(A∪B) = P(A ó B) = P(A) + P(B) - P(A y B)**

La probabilidad de que ocurra el evento A o el B es la suma de las probabilidades individuales menos la probabilidad de la intersección. Si dos eventos A y B, no tienen resultados en común (son disjuntos), entonces la probabilidad de que uno o el otro ocurra es simplemente la suma de las probabilidades individuales

Si A y B son **eventos disjuntos**, entonces P(A ó B) = P(A) + P(B)

Nota: Este caso especial puede extenderse a más de dos eventos. Si los eventos A, B y C **son disjuntos** entonces P(A ó B ó C) = P(A) + P(B) + P(C).

EJEMPLO 2. Género versus Educación

Volvamos a examinar los datos de la muestra aleatoria de 200 adultos clasificados según género y máximo nivel educativo alcanzado.

	Primario	Secundario	Terciario
Varón	38	28	22
Mujer	45	50	17

Considere los eventos:

A = "el adulto elegido tiene nivel terciario"

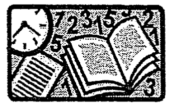
B = "el adulto elegido es varón y el máximo nivel educativo alcanzado es secundario"

C = "el adulto elegido es mujer"

¿Cuál es la probabilidad que un adulto elegido al azar haya alcanzado nivel terciario, o sea mujer?

$P(A \text{ ó } C) = P(A) + P(C) - P(A \text{ y } C) =$
$39/200 + 112/200 - 17/200 = 134/200 = 0.67$

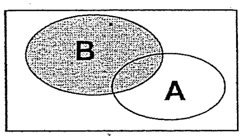
PARA RESOLVER !!!
6.12. Ganando contratos



Una compañía local de construcción entró en la licitación de dos contratos para la ciudad.
Ellos creen que la probabilidad de ganar el primer contrato es 0.6. La probabilidad de ganar el segundo contrato es 0.4. Por último, la probabilidad de ganar ambos contratos es 0.2.

(a) ¿Cuál es la probabilidad de que la compañía gane al menos uno de los dos contratos?

(b) Dibuje un diagrama de Venn que muestre los dos eventos
A = "ganar el primer contrato",
B = "ganar el segundo contrato".
Incluya todas las probabilidades de modo que la suma totalice 1.



(c) ¿Cuál es la probabilidad de ganar el primer contrato pero no el segundo?
Ayuda: sombree las partes del diagrama que representan el evento de interés.

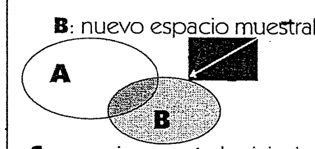
(d) ¿Cuál es la probabilidad de ganar el segundo contrato pero no el primero?

(e) ¿Cuál es la probabilidad de no ganar ningún contrato?

Algunas veces tendremos información conocida sobre el resultado de un experimento. Sería interesante actualizar la probabilidad de ocurrencia de cierto evento teniendo en cuenta la información conocida de antemano. Consideremos el experimento de tirar un dado correcto una sola vez. El espacio muestral de este experimento es S = {1,2,3,4,5,6} y cada uno de los seis resultados posibles es igualmente probable. La probabilidad de obtener "1" es 1/6. Pero supongamos que se sabe que el resultado fue un número impar, ¿cuál es ahora la probabilidad de "1"?

Dado que sabemos que el resultado fue un impar, no seguiremos considerando el espacio muestral original como el conjunto de resultados posibles. Hay sólo tres resultados posibles en el espacio muestral actualizado, {1,3,5}. Cada uno de estos tres resultados es igualmente posible, así que la probabilidad actualizada es 1/3. Lo que acabamos de calcular es una probabilidad condicional, la probabilidad de que haya ocurrido el evento A={1}, dado que ocurrió el evento B={impar}, representada por la expresión **P(A/B)**. En otras palabras, condicionados por el hecho de que el evento B ha ocurrido, deseamos encontrar la probabilidad actualizada de que ocurra el evento A.

A dado B



Nuestra próxima regla nos dice cómo hallar esas probabilidades condicionales. La base de esta regla es sencilla de comprender mirando el correspondiente diagrama. Como sabemos que el evento B ha ocurrido, tomaremos en cuenta sólo los resultados del suceso B. Este conjunto de resultados es nuestro espacio muestral actual y será la base para el cálculo de probabilidades. Deseamos encontrar la probabilidad de que ocurra el evento A en este nuevo espacio muestral. Los únicos resultados del evento A incluidos en este nuevo espacio muestral son aquellos que pertenecen tanto al evento A como al B (es decir, los resultados que pertenecen a la intersección).

Probabilidad Condicional

6.5. La probabilidad condicional del evento A dado que el evento B ya ha ocurrido:

$$P(A / B) = \frac{P(A \text{ y } B)}{P(B)}, \qquad \text{donde } P(B) > 0$$

Nota: Podemos volver a escribir la expresión anterior para calcular la probabilidad de una intersección, obtenemos la **regla de la multiplicación.**

P (ambos eventos ocurren al mismo tiempo) =

$$P(A \cap B) = P(A \text{ y } B) = P(A)\, P(B/A)$$

o

$$P(A \cap B) = P(A \text{ y } B) = P(B)\, P(A / B)$$

La base de esta regla es la siguiente: para que ambos eventos ocurran, primero debe ocurrir uno, por ejemplo el A, y luego de que ocurra el A, el evento B también debe ocurrir. Por supuesto los eventos A y B pueden cambiar sus lugares y entonces tenemos la segunda expresión.

EJEMPLO 3. Género versus Educación

Retomemos los datos de la muestra aleatoria de 200 adultos clasificados por género y mayor nivel educativo alcanzado.

	Primario	Secundario	Terciario
Varón	38	28	22
Mujer	45	50	17

Considere los eventos:
A = "el adulto elegido tiene nivel terciario"
C = "el adulto elegido es mujer"

¿Cuál es la probabilidad de que el adulto seleccionado al azar tenga nivel terciario **dado** que es mujer? Como sabemos que el adulto seleccionado es mujer, sólo consideraremos las 112 mujeres de nuestro espacio muestral actual. De esas 112 mujeres, hay 17 que tienen educación terciaria.

$$P(A / C) = P(\text{nivel terciario / mujer}) = \frac{17}{112} = 0.152$$

Si formalmente usamos la regla:

$$P(A / C) = \frac{P(A \text{ y } C)}{P(C)} = \frac{17/200}{112/200} = \frac{17}{112} = 0.152$$

Si la información sobre los eventos se presenta en forma de tabla de dos entradas, encontrar las probabilidades condicionales es bastante directo. En el ejemplo precedente, el suceso conocido fue "mujer", consideramos sólo la columna de las mujeres y expresamos la cantidad con educación terciaria como una fracción del total de mujeres.

☑ **PARA RESOLVER !!!**
6.13. Discos de computadora
Se les preguntó a 300 usuarios de PC qué tamaño y qué densidad de disco habían comprado. Los

resultados aparecen resumidos a la derecha.

	TAMAÑO	
DENSIDAD	5.25"	3.5"
Regular	80	60
Alta	60	100

(a) ¿Cuál es la probabilidad de que si se elige un usuario al azar compre disco de tamaño 3.5"?

(b) ¿Cuál es la probabilidad de que si se elige al azar un usuario de disco de alta densidad compre disco de tamaño 3.5"? Tenga en cuenta que una vez que seleccionó un usuario de disco de alta densidad, esto es información dada.

(c) Dé un ejemplo de dos eventos mutuamente excluyentes.

☑ **PARA RESOLVER !!!**
6.14. Probabilidades condicionales

Escenario I: Considere la experiencia de tirar un dado una sola vez. El espacio muestral es $S = \{1, 2, 3, 4, 5, 6\}$

(a) ¿Cuál es la probabilidad de tirar un 2? $P(2) =$

(b) Suponga que se sabe que el resultado anterior fue un impar, ¿cuál es ahora la probabilidad de tirar un 2?
$P(2 / \text{impar}) =$

Escenario II: Considere le experiencia de tirar dos veces una moneda.
El espacio muestral es $S =$ CC, CX, XC, XX .

(a) ¿Cuál es la probabilidad de obtener cara en la segunda tirada?
P (C en la segunda tirada) =

(b) ¿Cuál es la probabilidad de obtener cara en la segunda tirada si se obtuvo una cara en la primera?
P (C en la segunda tirada / C en la primera) =

PARA PENSAR !!!

Compare las respuestas de las partes **(a)** y **(b)** de los distintos escenarios del ejercicio anterior.

¿Cuáles son las diferencias entre estos dos escenarios?

Considere la siguiente situación: P (A / B) = 0.3 y P (A) = 0.3
¿Qué nos dice esto sobre los eventos A y B?
Esto es lo que ocurre cuando tiramos una moneda como en el ejercicio anterior.

Si saber que el evento B ha ocurrido no cambia la probabilidad de ocurrencia del evento A, es decir, P (A / B) = P (A), decimos que los dos eventos son **independientes**.

Definición
Dos eventos **A** y **B son independientes** si:
P (A / B) = P (A), o P (B / A) = P (B)

Si dos eventos no se influyen mutuamente (si saber que uno ha ocurrido no cambia la probabilidad de que ocurra el otro), los eventos son independientes. Si los dos eventos son independientes la regla general de la multiplicación dice que la probabilidad de que ambos eventos ocurran juntos es el producto de las probabilidades individuales.

Si dos eventos **A** y **B** son **independientes** ⇒ **P (A y B) = P (A) P (B)**

EJEMPLO 6.4. Una guirnalda de luces

Una guirnalda de luces contiene 30 bombitas. Si una de ellas falla, toda la guirnalda falla. La probabilidad de que una bombita funcione correctamente por lo menos 2 años es 0.98.

Si las bombitas funcionan independientemente, ¿cuál es la probabilidad de que la guirnalda de 30 luces funcione por lo menos 2 años?

P (guirnalda funcione) = P (las 30 luces funcionen)
= P (1ª funcione y 2ª funcione y... y 30ª funcione)

= P (1ª funcione) P(2ª funcione)... P (30ª funcione)
= (0.98) 30 = 0.545

Usamos la regla de la multiplicación para eventos independientes.

☑ **PARA RESOLVER !!!**
6.15. El plan familiar otra vez

Revisemos el PARA RESOLVER!!! 6.1, donde analizamos el caso de una pareja que va a tener hijos hasta tener un varón o tener 3 mujeres, lo que ocurra primero. Mediante la simulación estimamos la probabilidad de que la pareja tenga un varón. ¿Cuál fue esa estimación?

Ahora tenemos los conocimientos suficientes de cálculo de probabilidades para encontrar la verdadera probabilidad de 0.875. Suponemos que:

1. cada nacimiento tiene probabilidad 0.5 de ser varón y 0.5 de ser mujer,
2. el género de los nacimientos sucesivos es independiente, es decir, saber el género de un bebé no influye en el género de ninguno de los siguientes

S =		**Probabilidad**
	V →	$\frac{1}{2}$
	MV →	$\frac{1}{2}\frac{1}{2}$
	MMV →	$\frac{1}{2}\frac{1}{2}\frac{1}{2}$
	.... →	

Primero necesitamos generar el espacio muestral. A la derecha listamos tres de los cuatro eventos posibles. Agregue el faltante. A continuación calcule las probabilidades de cada resultado, usando los supuestos anteriores. Algunas de las probabilidades ya han sido calculadas. Encuentre la restante y verifique que la suma da 1.
Para encontrar la probabilidad de tener un varón se deben sumar las probabilidades de los eventos que corresponden a tener un varón, es decir, las probabilidades de los tres primeros resultados del espacio muestral.

P (tener un varón) = ...
Si convierte el resultado a forma decimal debe obtener 0.875.

☑ **PARA RESOLVER !!!**

6.16. ¿Buen negocio?

El *Good (not Better) Business Bureau* llevó a cabo una investigación sobre la calidad de los servicios ofrecidos por 86 talleres de reparación de autos de cierta ciudad. Los resultados en cuanto a **Servicio** y **Tipo de negocio** se resumen en la tabla siguiente.

	Servicio	
Tipo de Negocio	Bueno	Cuestionable
Agencia de Ventà de 0 km	18	6
Cuentapropista	34	28

(a) ¿Cuál es la probabilidad de que un negocio elegido al azar ofrezca buen servicio?

(b) ¿Cuál es la probabilidad de que un negocio elegido al azar sea un cuentapropista?

(c) ¿Cuál es la probabilidad de que un negocio elegido al azar sea cuentapropista y ofrezca buen servicio?

(d) ¿Cuál es la probabilidad de que un cuentapropista elegido al azar ofrezca buen servicio?

(e) Los eventos "cuentapropista" y "ofrece buen servicio" ¿son **mutuamente excluyentes**? Explique.

(f) Los eventos "cuentapropista" y "ofrece buen servicio" ¿son **independientes**? Explique.

EJERCICIOS

6.5. ¿Cuál es el error? Identifique el error, si es que lo hay, en cada afirmación. Si no hay error, escriba "Sin error".

a) Las probabilidades de que un vendedor de automóviles venda 0, 1, 2, o al menos 3 autos un día determinado de febrero son 0.19, 0.38, 0.29, y 0.15 respectivamente.

b) La probabilidad de que llueva mañana es 0.40, la probabilidad de que mañana no llueva es 0.52.

c) Las probabilidades de que una impresora cometa 0, 1, 2, 3, o al menos 4 errores al imprimir un documento son 0.19, 0.34, -0.25, 0.43, y 0.29 respectivamente.

d) Si se extrae una sola carta de un mazo de baraja de póquer, la probabilidad de seleccionar un corazón es 1/4, la probabilidad de seleccionar una carta negra (trébol o pique) es 1/2, y la probabilidad de seleccionar ambos a la vez (corazón y carta negra) es 1/8.

6.6. Una moneda y un dado: El proceso aleatorio consiste en tirar una moneda, si sale cara se la arroja por segunda vez. Si en la primer tirada sale cruz, se tira un dado una sola vez.

a) Escriba el espacio muestral del experimento.
b) Definamos los eventos A y B como sigue:
A = "cruz en la primera tirada, as al tirar el dado" = {X1}
B = "cara en la primera tirada y en la segunda" = {CC}
Encuentre la probabilidad de estos eventos.

6.7.

a) Si dos eventos son mutuamente excluyentes, sus complementos ¿también lo son? Explique

b) Si dos eventos son independientes, sus complementos ¿también lo son? Explique

6.8. Se clasificaron los doscientos adultos de una muestra de acuerdo al género y nivel educativo alcanzado. Si se elige una persona al azar de este grupo, ¿cuál es la probabilidad de que:

		Educación		
		Primaria	Secundaria	Terciaria
Género	Hombres	38	28	22
	Mujeres	45	50	17

a) sea hombre dado que tiene educación secundaria?
b) no tenga educación terciaria, dado que es una mujer?
c) Los eventos "hombre" y "educación secundaria" ¿son independientes? Presente los cálculos que avalen su respuesta

6.9. La compañía telefónica de una gran área metropolitana sabe que el

82% de sus clientes paga la factura bimestral en término. Suponga que se eligen al azar dos clientes del listado de todos los clientes y que las respuestas que darán son independientes. ¿Cuál es la probabilidad de que al menos uno pague su próxima factura en término?

6.10. Un artículo titulado "Controle su colesterol" (Fuente: Ann Arbor News, 8 de marzo de 1995) describe una prueba casera de colesterol elaborada por ChemTrack. El artículo dice que "los americanos se están volviendo más precavidos con el colesterol... 65% de los adultos mide su nivel de colesterol". Suponga que seleccionamos dos adultos americanos al azar. Considere que las respuestas de los dos son independientes. ¿Cuál es la probabilidad de que ambos hayan medido su nivel de colesterol?

6.11. En un experimento para estudiar la dependencia entre la hipertensión y el cigarrillo, se recogió la siguiente información sobre 180 personas.

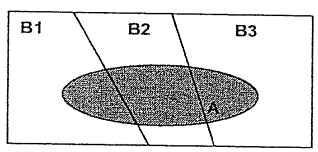
	No fumadores	Fumadores moderados	Fumadores empedernidos
Hipertensos	21	36	30
No hipertensos	48	26	19

a) Si se elige al azar un fumador empedernido de esta población, ¿cuál es la probabilidad de que sufra de hipertensión?

b) ¿Cuál es el porcentaje de fumadores?

c) Encuentre la distribución condicional del tipo de fumador para los hipertensos.

6.4.3 Partición y Teorema de Bayes



Deseamos encontrar la probabilidad de un evento A, pero hallarla no es siempre tan directo. Suponga que se tiene información sobre la ocurrencia del evento A, no en el espacio muestral como un todo, sino en varios subgrupos de dicho espacio muestral, por ejemplo B1, B2, y B3, como se muestra a la derecha. Nos gustaría combinar las probabilidades de A en cada subconjunto para obtener la P (A) total. Los tres eventos, B1, B2, y B3, forman una **partición** del espacio muestral.

Definición

Los eventos B1, B2, ..., Bl conforman una **partición** del espacio muestral S si:

1. B1, B2, ..., Bl son mutuamente excluyentes
2. La unión de B1, B2, ..., Bl es el espacio muestral **S**

Nota: Los eventos forman una partición si cada resultado individual del espacio muestral pertenece exactamente a uno de los eventos.

¿Cómo encontrar la probabilidad total del evento A? Primero considere las tres partes disjuntas del evento A:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$

Luego usamos la regla de la multiplicación para dos eventos y la aplicamos en cada una de las tres probabilidades del lado derecho.

$$P(A) = P(A / B_1)\,P(B_1) + P(A / B_2)\,P(B_2) + P(A / B_3)\,P(B_3)$$

Las tres probabilidades condicionales dadas para el evento A en cada parte del espacio muestral S se amalgaman ponderándolas por la probabilidad de estar en ese particular subconjunto de S. Esta última línea es la llamada **Regla de la Partición** o **Ley de la Probabilidad Total**.

Ley de la Probabilidad Total

6.6 Si los eventos B_1, B_2, B_3 forman una partición de S, entonces

$$P(A) = P(A / B_1)\,P(B_1) + P(A / B_2)\,P(B_2) + P(A / B_3)\,P(B_3)$$

Este resultado puede extenderse a una partición de más de tres eventos: Si los eventos B_1, B_2, ..., B_l forman una partición de S, entonces

$$P(A) = P(A / B_1)\,P(B_1) + P(A / B_2)\,P(B_2) + \ldots + P(A / B_l)\,P(B_l)$$

EJEMPLO 5. Un experimento en dos etapas

Suponga que tiene dos canastas que contienen pelotas de distintos colores:

Considere el siguiente experimento en dos etapas:

Canasta I	Canasta II
2 rojas 2 blancas	4 rojas 6 blancas

Etapa 1: elegir una canasta al azar, es decir, seleccionarla con igual probabilidad

Etapa 2: de la canasta elegida, tomar una pelota al azar

¿Cuál será la probabilidad de obtener una pelota roja?

$$P(\text{Roja}) = ?$$

Obviamente la probabilidad depende de la canasta seleccionada. Las dos canastas forman una partición.

a) ¿Cuál es la chance de obtener una pelota roja de la canasta I?

$$P(R / CI) = 2/4$$

b) ¿Cuál es la chance de obtener una pelota roja de la canasta II?

$$P(R / CII) = 4/10$$

Como en la primera etapa la canasta se elige al azar, la probabilidad de elegir una cualquiera es 1/2.

$$P(CI) = P(CII) = 1/2$$

Por último usamos la regla de la partición para combinar las dos probabilidades de a) y b).

$$\begin{aligned} P(\text{Roja}) &= P(\text{Roja}/ CI)\, P(CI) + P(\text{Roja}/ CII)\, P(CII) \\ &= (2/4)(1/2) + (4/10)(1/2) \\ &= 0.45 \end{aligned}$$

> ☑ **PARA RESOLVER !!!**
>
> **6.17. Las zapatillas equivocadas**
>
> Una compañía que fabrica calzado deportivo tiene dos fábricas. La fábrica A produce el 75% de las zapatillas y la B el restante 25%. Aproximadamente el 1% de las zapatillas de la fábrica A tiene mal puesto el número de calzado, mientras que el 2% de las de la fábrica B tiene este problema. Si usted compra un par de zapatillas de esta marca, ¿cuál es la probabilidad de que tenga mal puesta la numeración?

Una aplicación interesante del teorema de la partición es el **Modelo de Respuesta Aleatorizada de Warner**, que desarrollamos en el ejemplo siguiente.

Suponga que se desea estimar la proporción de estudiantes secundarios de la ciudad que usan droga. Preguntar directamente puede proporcionar información de escaso valor. ¿Cómo podemos obtener respuestas precisas a una pregunta tan ríspida que las personas son reacias a contestar la verdad?

EJEMPLO 6. Cómo obtener respuesta a preguntas delicadas

Método

Implemente una investigación con dos preguntas. Sea P1 la pregunta delicada. La segunda P2, es una pregunta cualquiera de la que se conozca la proporción de respuestas Sí que se espera obtener. Por ejemplo:

P1: "¿Alguna vez se ha llevado algo de un negocio sin pagarlo?" Sí NO

P2: "El segundero del reloj de la pared, ¿está entre el 12 y el 3?" Sí NO

Punto Clave

El encuestado determina cuál de las preguntas responde usando algún método probabilístico que está bajo su control. Por ejemplo, el encuestado puede tirar una moneda. Si sale cara responde pregunta P1, si sale cruz responde la pregunta P2. Como sólo el que responde sabe cuál es la pregunta que contesta, no debería sentir temor o vergüenza por su respuesta.

Pero, ¿podemos aún encontrar la proporción de los que responden Sí a la pregunta delicada? Queremos saber sobre la proporción de la población para los que la respuesta a la pregunta Q1 es SI, esto es la probabilidad de un "Sí" a la pregunta P1. Considerando una partición generada por las preguntas P1 y P2 y aplicando la Regla de la Partición o Ley de la Probabilidad Total:

$$P(SI) = P(SI / P1)\, P(P1) + P(SI / P2)\, P(P2)$$

Por la experiencia aleatoria usada para determinar cuál pregunta se responde, conocemos los valores de P (P1) y P (P2). Conocemos también el valor de P (SI / P2), para la pregunta del segundero del reloj de la pared, la probabilidad es 0.25. Lo que resta es la probabilidad de interés, P (SI / P1), que se puede resolver.

Por ejemplo, si el 37.5% de los estudiantes responde "Sí" entonces:

$$0.375 = P(SI / P1)\,(0.5) + (0.25)\,(0.5)$$

$$\Rightarrow P(SI / P1) = [\,0.375 - (0.25)\,(0.5)]\,/\,(0.5) = 0.50$$

Estimamos que la proporción de estudiantes que alguna vez se llevaron algo de un negocio sin pagar es 50%.

☑ **PARA RESOLVER !!!**

6.18. Obtener repuesta a preguntas delicadas

Realice el Modelo de Respuesta Aleatorizada de Warner con sus compañeros de clase.

- Determine una pregunta delicada de interés y regístrela como pregunta P1.
- Seleccione dos mecanismos aleatorios para el modelo -uno para determinar cuál de las dos preguntas se debe contestar y el otro para la segunda pregunta, P2 en la investigación- para esta última se debe conocer la proporción de veces que se espera una respuesta "Sí"

Investigación

P1: ¿.......................................? Sí NO

P2: Al tirar una moneda por segunda vez, ¿obtuvo una cara? Sí NO

Procedimiento

Distribuya tiras de papel, una por cada estudiante, para que anote su respuesta. Sólo se deberá escribir una palabra en la tira de papel, la respuesta del estudiante a la pregunta que haya resultado sorteada. No se deben escribir nombres en el papel, ni tampoco la pregunta que se respondió. (Por ejemplo, no escriba "P1 = No").
Recoja los resultados y determine la proporción de respuestas "Sí".

Introduzca los valores en la siguiente expresión y estime P(SI / P1)

$$P(SI) = P(SI / P1)\,P(P1) + P(SI / P2)\,P(P2)$$

La Regla de la Partición también juega un papel importante en el ejemplo siguiente, para determinar la tasa de falsos positivos de una prueba de diagnóstico, es decir la proporción de gente que no tiene la enfermedad pese a que la prueba dio resultado positivo.

EJEMPLO 7. Prueba para una enfermedad rara

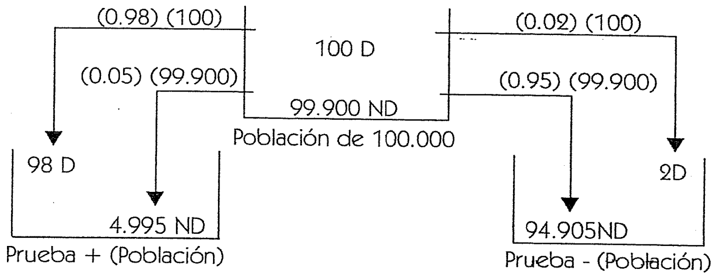
Suponga que se ha desarrollado una prueba muy confiable para una enfermedad rara como el SIDA o la hepatitis. En particular, suponga que cuando la enfermedad está presente, la prueba da positiva el 98% de las veces. Aplicada a personas sanas da negativa el 95% de las veces. Los términos médicos para estos dos porcentajes son especificidad y sensibilidad, respectivamente. También suponga que el 0.1% de la población padece la enfermedad.

Si la prueba no implica costos altos para su implementación, ¿se debería llevar a cabo una investigación pública de gran escala? Con los números precedentes, usted ¿se preocuparía si resultara positivo en la investigación anterior?

¿Se sorprendería si, sin considerar más información que la anterior, anticipamos que la probabilidad de realmente tener la enfermedad cuando la prueba dio positiva es sólo de alrededor del 2%?

Considere la siguiente notación y el diagrama, que muestra los resultados para 100.000 personas.

Sea **D** el evento "tener la enfermedad"
ND el evento "no tener la enfermedad"
+ el evento "la prueba dio positiva"
- el evento "la prueba dio negativa"



• De las 100.000 personas elegidas aleatoriamente de la población, se espera que el 0.1%, o sea, 100 personas tengan la enfermedad y las restantes 99.900 no la tengan.

• De las 100 personas que padecen la enfermedad, se espera que al 98%, o sea, 98 personas la prueba les dé positiva, mientras que a las 2 restantes le dará negativa.

• De las 99.900 personas que no tienen la enfermedad, se espera que al 95%, o sea, 94.905 personas (= 99.900*0.95) la prueba le dé negativa y a las restantes 4.995 personas positiva.

• Así, de las 5.093 personas (4.995+98) a quienes la prueba les dio positiva, sólo 98 (aproximadamente el 2%) tienen la enfermedad.

En las personas positivas siempre se requieren pruebas confirmatorias. Considere también el impacto social y psicológico de recibir un resultado positivo si uno no está familiarizado con las estadísticas. O sea, las investigaciones públicas de gran escala para una enfermedad realmente rara generalmente son una mala idea. Sin embargo, la clave es que la enfermedad sea verdaderamente rara. Si la incidencia de la enfermedad fuera 10% en lugar de 0.1%, entonces la probabilidad de tener la enfermedad si la prueba dio positiva, es aproximadamente 69%, en lugar del 2%. Compruebe este valor con un gráfico como el anterior.

☑ **PARA RESOLVER !!!**

6.19. Test diagnóstico para el SIDA

Apliquemos todo lo que aprendimos de probabilidades en el siguiente ejemplo concerniente a un análisis de sangre para detectar la presencia de virus del SIDA. Considere la notación usada en el ejemplo previo:

D =	Tiene la enfermedad	**+** =	El test es positivo
ND =	No tiene la enfermedad	**-** =	El test es negativo

Las siguientes son propiedades de un test para estudiar la presencia de virus del SIDA en sangre:

(1) Cuando los anticuerpos están presentes en sangre, el test da resultado positivo con probabilidad 0.98.

P (+ D) = 0.98

(2) Cuando los anticuerpos no están presentes en sangre, el test da resultado negativo con probabilidad 0.95.

P (- ND) = 0.95

(3) Suponga que el 0.1% de la población tiene anticuerpos del SIDA

P (D) = 0.001

1. ¿Cuál es la probabilidad a priori que una persona elegida al azar de la población, tenga la enfermedad? Esta es justamente la propiedad (3) de más arriba.

P (D) =

2. Si se le hace el test a todas las personas de la población, ¿qué proporción de los tests dará positivo?

P (+) =

3. Se elige una persona al azar, se le practica el test y resulta positivo, ¿cuál es la probabilidad de que tenga anticuerpos de SIDA? A este valor se lo llama también probabilidad posterior.

P (D / +) =

Al complemento de la probabilidad del punto 3 se lo llama también tasa de falsos positivos:

P (ND / +) = 1 - P (D / +) =   ¿Por qué es tan alta?

Teorema de BAYES

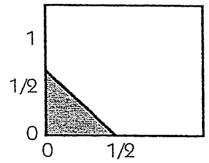
6.7. Suponga que los eventos B_1 y B_2 conforman una partición (son sucesos complementarios cuyas probabilidades suman 1). Suponga que A es otro evento y que conocemos las probabilidades condicionales de $P(A/B_1)$ y $P(A/B_2)$. Entonces la **regla de Bayes** nos proporciona otra probabilidad condicional:

$$P(B_1/A) = \frac{P(A/B_1)\,P(B_1)}{P(A/B_1)\,P(B_1) + P(A/B_2)\,P(B_2)}$$

EJEMPLO 8. Probabilidad Geométrica

Considere el siguiente cuadrado de superficie uno:

Si se arroja un dado al azar, ¿cuál es la probabilidad de que caiga en la zona sombreada? Suponga que P(el dardo dé en el cuadrado) = 1

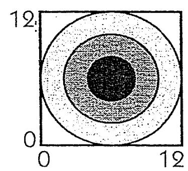


Volvamos a la definición frecuentista de la probabilidad, proporción o frecuencia relativa en gran cantidad de repeticiones. Dado que el dardo se arroja al azar la probabilidad de acertar en la región sombreada es igual al cociente entre el área de la región sombreada y el área total del cuadrado. Recordemos también que el área del triángulo es la mitad de la base por la altura.

$$P\text{ (región sombreada)} = \frac{\text{Área de la región sombreada}}{\text{Área total}} = \frac{\frac{1}{2}\,\frac{1}{2}\,\frac{1}{2}}{1} = 1/8 = 0.125$$

☑ PARA RESOLVER !!!

6.20. Tiro al blanco



El propósito del nuevo juego con dardos es acertar tan cerca del centro del objetivo como sea posible.

➤ Si se impacta dentro del anillo #1, el más oscuro, se gana el gran premio de $10.

➤ Si se impacta dentro del anillo #2, pero fuera del #1, se obtiene el segundo premio: un animal grande embalsamado.

➤ Si se impacta dentro del anillo #3, pero fuera del #2, se obtiene el tercer premio: un animal pequeño embalsamado.

➤ Si se impacta fuera del anillo #3, no se gana nada.

El juego tiene un mecanismo que dispara el dardo hacia el objetivo al azar cuando se presiona un botón.

Suponiendo que el dardo se arroja al azar, y

- P (dardo caiga en el cuadrado) = 1,
- el radio del anillo #1 es 2,
- el radio del anillo #2 es 4,
- el radio del anillo #3 es 6.

Y recordando que el área del círculo es: Área = r2

¿Cuál es la probabilidad de:

(a) ganar el gran premio? P = 4 / 144 = 0.08

(b) ganar el segundo premio?

(c) ganar el tercer premio?

(d) no ganar nada?

Los dueños piensan que muchos jugadores van a ganar el primer premio y desean hacer más pequeño el radio del anillo #1. ¿Cuál debe ser el radio para que la probabilidad de ganar el gran premio sea sólo de 0.01?

EJERCICIOS

6.12. Se llevó a cabo una investigación para estudiar la presencia de bacterias en una población de animales. Para hacer el diagnóstico se usó una herramienta con las siguientes propiedades:
Cuando se aplica en animales infectados, 98% de los tests indican infección. Cuando se aplica a animales que no están infectados, 95% de los tests indican ausencia de infección. Suponga que el 2% de los animales están infectados.

a) ¿Qué proporción de los animales que no están infectados presentan test con indicación de infección?

b) De los animales cuyos tests indican infección, ¿qué porcentaje está realmente infectado?

c) De los tests que indican infección, ¿qué porcentaje corresponde a no infectados?

d) Si la población en estudio correspondiera a un grupo de alto riesgo con 20% de realmente infectados, en lugar de 2%, ¿la respuesta b) aumentaría, disminuiría o seguiría igual? Explique.

6.13. Una caja contiene 5 bolitas rojas y diez blancas. Se extraen dos al azar sin reposición.

a) ¿Cuál es la probabilidad de que la segunda bolita sea roja?

b) ¿Cuál es la probabilidad de que la segunda sea roja, dado que la primera resultó blanca?

6.14. Hay tres comisiones de Estadística. El 38% de los alumnos asiste a la comisión A, el 30% a la B y el resto a la C. El 95% de los estudiantes de la comisión A entiende la regla de Bayes, en la comisión B la entendió el 84% y en la C sólo la comprendió el 8% de los estudiantes. Si un alumno elegido al azar dice entender la regla, ¿cuál es la probabilidad de que pertenezca a la comisión A?

6.15. Dos corporaciones de software, la Corp I y la Corp II, compiten para tomar el control de una tercera corporación. Se estima que la probabilidad de que la Corp I sea la exitosa es 0.7, y la probabilidad de que triunfe la Corp II es 0.3. Si la Corp I logra el control la probabilidad de que el año próximo desarrolle un nuevo producto es del 0.8. Si triunfa la Corp II la probabilidad de desarrollar el nuevo producto el próximo año es sólo 0.4.

a) ¿Cuál es la probabilidad de que el año próximo se desarrolle el nuevo producto?

b) Suponga que el año próximo se desarrolla el nuevo producto, ¿cuál es la probabilidad de que sea la Corp I la que tomó el control?

6.5. VARIABLES ALEATORIAS

En los capítulos anteriores desarrollamos distintos tipos de variables y modos de resumir la distribución de esas variables. Usamos modelos para describir la distribución y para encontrar la proporción de unidades que tienen ciertos valores. En esta sección mezclaremos las ideas sobre distribución de variables con definiciones más formales de probabilidad e introduciremos la idea de variable aleatoria y su correspondiente distribución de probabilidad.

Cuando escribimos algunos de los espacios muestrales de varios experimentos observamos que el espacio muestral no necesariamente debe ser una serie de números, aunque se puede usar un código cuando el resultado no es numérico. A los estadísticos les gusta tratar con resultados numéricos, y eso nos conduce a la próxima definición.

> **Definición**
> Una **variable aleatoria** es una función real definida en un espacio muestral.

Podemos pensarla como una regla que asigna un valor numérico (y sólo uno) a cada punto del espacio muestral generado por un proceso aleatorio. En resumen, una variable aleatoria es el resultado numérico de un proceso aleatorio.

A las variables aleatorias usualmente se las denota con letras mayúsculas del final del alfabeto (...X, Y, Z)

EJEMPLO 9. Variables aleatorias

Nº 1. Si el experimento es arrojar una moneda tres veces, entonces la variable aleatoria X se podría definir como el número de caras en las tres tiradas.

X = # caras en las tres tiradas de una moneda
Los valores posibles de X serían 0, 1, 2, ó 3.

Nº 2. Si el experimento es elegir un estudiante al azar de la población de la universidad, entonces la variable aleatoria podría ser Y el número de horas de crédito del alumno para el período actual.

Y = # horas de crédito del período actual
Los posibles valores de Y podrían estar representados como 0, 1, 2,..., 20. El máximo de 20 horas podría ser el límite superior fijado por la universidad.

Nº 3. En la experiencia de seleccionar una lámpara al azar de una línea de producción, la variable aleatoria Z podría corresponder con el tiempo de duración de la lámpara medido en horas.

Z = tiempo de duración de la lámpara, en horas

Los valores posibles de Z se representan como el intervalo $[0, \infty]$, esto es, cualquier valor mayor o igual a 0. Ciertamente debe haber algún límite superior máximo razonable. Sin contar con más información no sabremos qué valor usar.

Del mismo modo que en el capítulo 4 diferenciamos variables discretas y continuas, también haremos la distinción entre variables aleatorias discretas y continuas. Una variable aleatoria que toma un número finito o infinito numerable de posibles valores se dice que es discreta. Una variable aleatoria se dice continua si la serie de sus valores posibles es un intervalo o un conjunto de intervalos de números reales. En el ejemplo 8.9, las variables aleatorias X, número de caras e Y, número de horas de crédito, son discretas, mientras que la variable aleatoria Z, que representa la duración de las lámparas, es continua.

> **Definición**
> Una **variable aleatoria discreta** puede tomar un número finito de valores o infinito numerable.
> Una **variable aleatoria continua** puede tomar cualquier valor de un intervalo o conjunto de intervalos.

Asociada a cada variable aleatoria hay un modelo –la función de probabilidad, si es discreta (p(x)) o la función de densidad de probabilidad (f(x)), si es continua. La diferencia más importante entre los modelos presentados en este capítulo y aquellos del capítulo 5 es que antes hablábamos en términos de "proporción" y ahora usaremos la palabra "probabilidad".

Variable aleatoria X
- Discreta → Función de probabilidad $p(x) = P(X=x)$
- Continua → Función de densidad de probabilidad: f(x) área bajo f (x) entre a y b = $P(a<X<b)$

6.5.1. Variables aleatorias discretas

Una variable aleatoria X que tiene un número finito o infinito numerable de valores posibles se llama **discreta**. Se asigna una probabilidad a cada uno de los posibles valores, donde cada probabilidad debe ser un número entre 0 y 1 y la suma de todas las probabilidades debe ser igual a 1. La **función de probabilidad de X**, será simbolizada con p(x) [llamada también función de probabilidad puntual o función de cuantía o de masa de probabilidad].

Si X es una variable aleatoria discreta que toma los valores $x_1, x_2,\ldots,x_k$, con probabilidades $p_1, p_2, \ldots, p_k$, entonces la **distribución de probabilidad de X**, se puede representar por una fórmula, una tabla o una gráfica

que indique las probabilidades p(x) correspondientes a cada uno de los valores de X; es decir:

X = x	x_1	x_2	...	x_k
P (X = x)	p_1	p_2	...	p_k

o bien: $$p(x) = \begin{cases} p_1 & \text{si } x = x_1 \\ p_2 & \text{si } x = x_2 \\ \vdots & \\ p_k & \text{si } x = x_k \end{cases}$$

> La **función de probabilidad de una variable aleatoria discreta X** es una función, a menudo indicada con p (x)= P(X=x), que satisface:
>
> **1)** $p(x) \geq 0$ para todo x **y**
> **2)** la suma de todos los p(x) es 1.

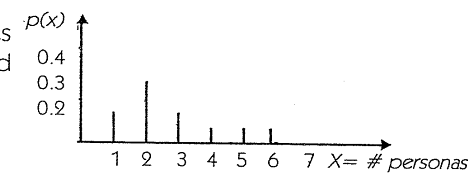
EJEMPLO 10. Número de personas por vivienda

Sea X el número de personas por vivienda en cierta comuni-dad, con la siguiente distribución de probabilidad:

X =x	1	2	3	4	5	6	7
P (X=x)	0.20	0.32	0.18	0.15	0.07	0.03	

a) ¿Cuál debe ser la probabilidad de encontrar 7 personas en una vivienda para que ésta sea una distribución de probabilidad de variable discreta?

Como las probabilidades deben sumar 1, la probabilidad de 7 personas debe ser:



$$1 - (0.20+0.32+0.18+0.15+0.07+0.03) = 0.05$$

b) Grafique la distribución de probabilidad. Se puede usar un "gráfico de bastones"

c) ¿Cuál es la probabilidad de que en una vivienda elegida al azar vivan más de 5 personas?

$P(X > 5) = 0.03 + 0.05 = 0.08$

d) ¿Cuál es la probabilidad de que en un vivienda elegida al azar vivan no más de 2 personas?

$P(X \leq 2) = 0.32 + 0.20 = 0.52$

e) ¿Cuál es $P(2 < X \leq 4)$: la probabilidad de que en una vivienda elegida al azar habiten más de 2 personas pero no más de 4?

$P(2 < X \leq 4) = 0.18 + 0.15 = 0.33$

☑ **PARA RESOLVER !!!**

6.21. Suma de puntos

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

Considere el experimento de arrojar dos dados y defina la variable aleatoria X como la suma de los valores de los dos dados. Antes encontró el espacio muestral correspondiente. Recuerde los 36 resultados posibles.

a) Escriba la distribución de frecuencias de X, es decir la lista de valores posibles de X y sus respectivas probabilidades. Luego represéntela con un gráfico de bastones, en el eje vertical indique las probabilidades.

Valor de X $X = x$	
Probabilidad $P(X=x)$	

b) Encuentre $P(\text{Suma} > 7)$

c) ¿Cuál es la probabilidad de obtener 7 u 11 en la próxima tirada de dos dados?

d) ¿Cuál es la probabilidad de obtener una suma de al menos 3 en la próxima tirada de dos dados?

☑ **PARA RESOLVER !!!**

6.22. Pareja #1 y Pareja #2

Cada una de dos parejas, llamémoslas Pareja #1 y Pareja #2, ha decidido tener exactamente tres hijos. Un resultado posible de cualquiera de las dos parejas es el evento "MMV", que significa que los dos primeros nacimientos corresponden a mujeres y el tercero es un varón.

a) Escriba el espacio muestral de la Pareja #1

$S = \{$

b) Sea X la variable aleatoria "número de mujeres de la Pareja #1". ¿Cuáles son los posibles valores de X?

Valores posibles = ...

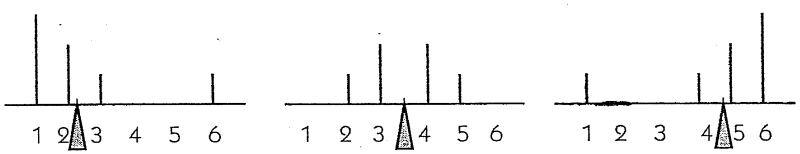
c) Suponga que cada nacimiento tiene probabilidad 0.5 de ser mujer y 0.5 de ser varón; y que el género de los nacimientos sucesivos es independiente. Dé la distribución de probabilidad de la variable aleatoria X: "número de mujeres de la Pareja #1"

Valor de X $X = x$	
Probabilidad $P(X=x)$	

d) Suponiendo que los resultados de la Pareja #1 son independientes de los resultados de la Pareja #2, ¿cuál es la probabilidad de que las dos parejas tengan el mismo número de mujeres?

> **RECORDAR !!!**
> Definición y propiedades de función de distribución de una variable aleatoria: $F(x) = P[X \leq x]$

En el capítulo 5 tratamos cómo resumir una serie de datos con medidas de posición y de dispersión. Antes de pasar a variables aleatorias continuas, aprenderemos a resumir una distribución de probabilidad discreta con la media y el desvío estándar. La **media de una variable aleatoria discreta X** es el punto de equilibrio del gráfico de bastones.

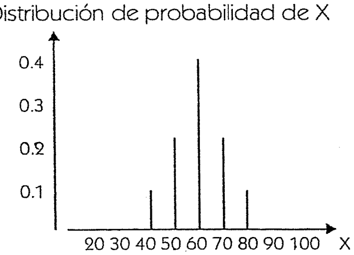


La media también llamada **valor esperado de X** se calcula como sigue:

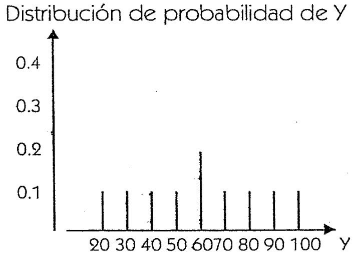
> **Definición**
> Si X es una variable aleatoria discreta que toma los valores $x_1, x_2, \ldots, x_k$, con probabilidades $p_1, p_2, \ldots, p_k$, entonces la **media**, o **valor esperado de X** [E(X)] viene dada por:
>
> $$E(X) = \mu_x = x_1p_1 + x_2p_2 + \ldots + x_kp_k = \sum_{i=1}^{k} x_i\,p_i$$

La **varianza** o **variancia** de una variable discreta X [σ^2] es una medida de la dispersión de los posibles valores respecto de la media. Examinando las dos distribuciones que se muestran debajo, podemos ver que, aun cuando la media de X es la misma que la de Y (ambas valen 60) la varianza de X es menor que la varianza de Y.

Distribución de probabilidad de X



Distribución de probabilidad de Y



La expresión para la varianza de X aparece debajo. La expresión del lado derecho de la igualdad, que es equivalente a la del lado izquierdo, se usa a los efectos del cálculo, como se muestra en el Ejemplo 8.13 (está corregido en el O. de A.: no se entiende). Si sacamos la raíz cuadrada de la varianza, tenemos el desvío estándar de X, burdamente interpretado como la distancia promedio de los valores posibles de X respecto de su media.

> **Definición**
> Si X es una variable aleatoria discreta que toma los valores $x_1, x_2, \ldots, x_k$, con probabilidades $p_1, p_2, \ldots, p_k$, entonces la *varianza*, de **X** viene dada por:
>
> $$\mathrm{Var}(X) = \sigma^2_x = E[(X - \mu)^2] = E(X^2) - [E(X)]^2$$
>
> y el **desvío estándar de X** viene dado por:
>
> $$SD(X) = \sigma_x = \sqrt{\sigma^2 x}$$

EJEMPLO 11. Otra vez la cantidad de personas por vivienda

X =x	1	2	3	4	5	6	7
P (X=x)	0.20	0.32	0.18	0.15	0.07	0.03	0.05

Basándonos en este modelo, ¿cuál es el número esperado de personas por vivienda?

$E(X) = \mu x = 1*0.20+2*0.32+3*0.18+4*0.15+5*0.07+6*0.03+7*0.05 = 2.86$

O sea, uno espera, en promedio, 2.86 personas por vivienda.

Para encontrar la **varianza** del número de personas por vivienda, primero es necesario calcular $E(X^2)$ según la forma básica de cálculo de esperanzas.

$E(X^2) = \sum_{i=1}^{7} x_i^2\,p_i =$

$= (1)^2(0.20)+(2)^2(0.32)+(3)^2(0.18)+(4)^2(0.15)+(5)^2(0.07)+(6)^2(0.03)+(7)^2(0.05) =$
$= 10.78$

Ahora tenemos,

$$Var(X) = \sigma^2 x = E(X^2) - [E(X)]^2 = 1078 - (2.86)^2 = 2.6$$

El desvío estándar es $SD(X) = \sigma_x = \sqrt{2.6} = 1.61$

Si se elige una vivienda al azar de esta comunidad, se espera que la cantidad de personas que viven en ella sea de aproximadamente 2.86, con una dispersión de 1.61.

☑ **PARA RESOLVER !!!**

6.23. Otra vez la suma de puntos

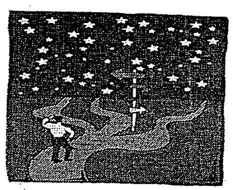
Considere la experiencia de arrojar dos dados. Sea X la variable correspondiente a la suma de los puntos obtenidos. La distribución de probabilidad aparece debajo.

Valor de X X = x	2	3	4	5	6	7	8	9	10	11	12
Probabilidad P (X=x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Calcule la media de X, es decir, el valor esperado de la suma de los puntos de los dos dados.
¿Es este valor consistente con la idea de ser el punto de equilibrio de la gráfica de bastones?

☑ **PARA RESOLVER !!!**

6.24. Los beneficios y el clima



Un promotor de espectáculos debe decidir la contratación de un concierto para el próximo otoño. El concierto puede realizarse en un auditorio cubierto o en uno al aire libre que es mucho más grande. Los siguientes son los pronósticos de beneficios (en cientos de dólares), para cada ubicación según las distintas condiciones climáticas.

CLIMA	Beneficios (cientos de $) Auditorio cubierto	Aire libre	PROBABILIDAD
Soleado y cálido	50	80	0.5
Soleado y frío	50	60	0.2
Lluvioso y cálido	60	40	0.2
Lluvioso y frío	40	6	0.1

(a) El beneficio esperado para el auditorio cubierto es $51.000.
Calcule el pago esperado asociado al auditorio al aire libre.

(b) Si el objetivo es maximizar el beneficio esperado, ¿cuál es la mejor ubicación?

(c) Si el promotor DEBE asegurar un beneficio de por lo menos $40.000, ¿cuál es la mejor ubicación?

(d) ¿Cuál es la ubicación que tiene la menor variabilidad en los beneficios? Explique.

6.5.2 Variables aleatorias continuas

Una variable aleatoria X que toma valores en algún intervalo, o unión de intervalos, se dice **continua**. No se puede asignar una probabilidad a cada valor posible ya que los valores posibles de un intervalo no se pueden contar. Asignamos una probabilidad 0 a cada valor individual o la suma de las probabilidades sería eventualmente mayor que 1.

La idea es asignar probabilidades a intervalos de resultados, no a valores individuales, y representar esas probabilidades como áreas debajo de una curva, llamada **curva de densidad de probabilidad**.

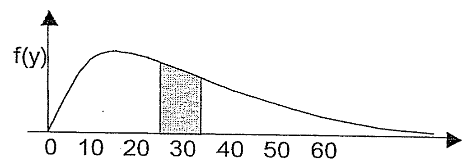
Definición

El comportamiento de una variable aleatoria continua **X** viene dado por una función, identificada con **f(x)**, tal que P (X tome valores en un subconjunto A) = el área bajo la función f (x) sobre el subconjunto A.
La función **f(x)** debe cumplir:
(1) $f(x) \geq 0$ para todo x **y** **(2)** el área total bajo f(x) es 1.

Esto es, la función **f(x)** debe ser una curva de densidad. Ya vimos curvas de densidad con anterioridad. Ahora, en lugar de referirnos a calcular áreas para encontrar proporciones, diremos que calculamos áreas para encontrar probabilidades.

Suponga que la función de densidad de probabilidad de Y = ingreso anual (en miles de dólares) de un adulto americano seleccionado aleatoriamente está dada por:

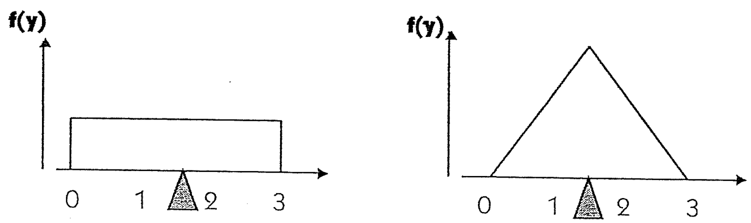
¿Cómo encontrar la probabilidad de seleccionar una persona al azar de esta población con un ingreso entre \$30.000 y \$40.000?



$P(30 \leq X \leq 40) =$

Use la función de densidad para encontrar el área entre $y=30$ e $y=40$.

El cálculo de la media μ y la varianza σ^2 cuando se dispone de una función de densidad de probabilidad es similar al de las expresiones del caso discreto, excepto por el hecho de que se necesita usar una rama del Cálculo llamada integración en lugar de la sumatoria. Sin embargo, la ubicación del centro de gravedad de la media de la densidad es evidente cuando se mira la gráfica de la función.



Definición
La **media de una variable aleatoria continua X** es el punto de equilibrio de la función de densidad de probabilidad

EJEMPLO 12. Un largo embarazo

Sea X la duración de un embarazo en días. X es una variable aleatoria continua. Suponga que tiene una distribución aproximadamente normal con promedio 266 días y desvío estándar 16 días.

¿Cuál es la probabilidad de que un embarazo dure al menos 310 días?

Queremos calcular $P(X > 310)$. Haremos los mismos cálculos que cuando trabajamos con la distribución normal pero ahora lo llamaremos probabilidad en lugar de proporción.

$$P(X > 310) = P\left(\frac{X-266}{16} > \frac{310-266}{16}\right) = P(Z > 2.75) =$$

$$= 1 - 0.9970 = 0.0030$$

☑ **PARA RESOLVER !!!**
6.25. Cáncer de piel

En los últimos años, el porcentaje de gente que contrae melanoma, un tipo de cáncer de piel, durante el verano en los estados del sur de los EEUU es una variable aleatoria que sigue una distribución uniforme entre 1.0% y 4.5%.

(a) Dibuje la distribución de X: porcentaje de gente que contrae melanoma.
(b) ¿Cuál es el porcentaje medio de gente que contrae melanoma?

(c) Encuentre la probabilidad de que en un verano determinado el porcentaje de gente que contrae melanoma sea menor que el 1.5%.

(d) Encuentre la probabilidad de que en un verano determinado el porcentaje de gente que contrae melanoma sea menor que el 1.5%, dado que es mayor que 1.2%.

EJERCICIOS

6.16. Alfredo está en un casino y quiere participar de un juego que tiene dos opciones.

Resultado	C	X
Ganancia neta	\$2	\$-2
Probabilidad	0.5	0.5

Ganancia neta esperada=0

Opción 1: Se arroja una vez una moneda correcta. El jugador gana \$4 si sale cara, nada en otro caso. Para jugar hay que pagar \$2.

Opción 2: Se arroja dos veces una moneda correcta. El jugador gana \$8 si sale cara en ambas tiradas, nada en otro caso. Para jugar se paga \$3.

Usando el criterio de la ganancia neta esperada ¿cuál de las dos opciones es más atractiva para Alfredo? Explique todos los detalles que apoyan su decisión.

6.17. Un caso interesante. Una variable aleatoria X toma los valores 1 y -1 con probabilidad 2/3 y 1/3 respectivamente. Sea Y = X2

a) ¿Cuál es el valor esperado de Y? E (Y) =

b) ¿Cuál es la varianza de Y? Var (Y) =

6.18. Lavadero de autos. Suponga que el número de autos X, que pasa por un lavadero automático de autos entre las 4 p.m. y las 5 p.m. de un viernes soleado tiene la siguiente distribución de probabilidad:

X = x	4	5	6	7	8	9
P(X = x)	1/12	1/12	1/4	1/4	1/6	

a) Complete la distribución

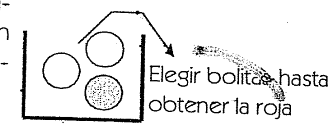
b) ¿Cuál es la probabilidad de que al menos 6 autos pasen por el lavadero entre las 4 p.m. y las 5 p.m. de cualquier viernes soleado?

c) ¿Cuál es E (X), el número esperado de autos que pasan por el lavadero entre las 4 p.m. y las 5 p.m. de cualquier viernes soleado?

d) En el capítulo 5 estudiamos transformaciones lineales. Vimos que la media de una nueva variable Y, que es una transformación lineal de la variable X, se encuentra introduciendo la media de X directamente dentro de la transformación. Si Y = 2X - 1 representa la cantidad de

dinero que el dueño del lavadero paga por hora al empleado que atiende, ¿cuál es el salario promedio que el empleado recibe en ese período en particular?

6.19. Con versus sin reemplazo. Una urna contiene 3 bolitas de las cuales una es roja y dos son blancas. El experimento consiste en extraer una bolita por vez sin reemplazo hasta obtener la bolita roja, y entonces detenerse.



(a) Escriba el espacio muestral de esta experiencia.

(b) Sea la variable aleatoria X el número de extracciones hasta que sale la bolita roja.

(c) X = número total de bolitas extraídas. ¿Cuáles son los valores posibles de esta variable aleatoria?

(d) Repita las partes (a) y (b) suponiendo que se trabaja con reemplazo.

RESUMEN DEL CAPÍTULO

Como extraemos muestras de las poblaciones, nuestras conclusiones o inferencias sobre la población tienen cierto grado de incertidumbre. Descubrimos que las probabilidades se pueden encontrar por simulación o como resultados matemáticos más formales. La simulación es una técnica poderosa cuando el planteo matemático del problema es difícil.

La probabilidad es una herramienta importante en los procesos de toma de decisión. Es necesario saber llevar adelante los cálculos de valor-p asociado a un resultado observado. En los capítulos siguientes veremos cómo utilizar este valor para el desarrollo formal de pruebas de hipótesis para la media de la población.

REGLAS BÁSICAS DE PROBABILIDAD

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

2. $P(S) = 1$

3. Regla del Complemento

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

4. Regla General de la Suma

$P(A \cup B) = P(A \text{ ó } B) = P(A) + P(B) - P(A \text{ y } B)$

5. Probabilidad condicional del evento B dado el evento A

$$P(B/A) = \frac{P(A \text{ y } B)}{P(A)}$$

Regla General de la Multiplicación

$P(A \cap B) = P(A \text{ y } B) = P(A)\,P(B/A) = P(B)\,P(A/B)$

6. Regla de la Partición

$P(A) = P(A/B_1)\,P(B_1) + P(A/B_2)\,P(B_2) + \ldots + P(A/B_l)\,P(B_l)$

7. Regla de Bayes

$$P(B1/A) = \frac{P(A/B1)\,P(B1)}{P(A)} = \frac{P(A/B1)\,P(B1)}{P(A/B1)\,P(B1) + \ldots + P(A/Bl)\,P(Bl)}$$

Palabras Clave

Asegúrese que es capaz de describir, en sus propias palabras, y dar un ejemplo de las siguientes palabras clave de este capítulo.

probabilidad
frecuencia relativa
proceso aleatorio
simulación
espacio muestral
eventos
diagrama de Venn
unión
intersección
complemento
disjunto (o mutuamente excluyentes)
equiprobable
regla del complemento
regla de la suma
probabilidad condicional
regla de la multiplicación
independencia
partición
regla de la partición
regla de Bayes
variable aleatoria (v a)
función de probabilidad
función de densidad
función de distribución
valor esperado (o promedio)
varianza de una v a

EJERCICIOS

6.20. Varivax, una nueva vacuna para la varicela, fue recientemente aprobada por la Food and Drug Administration de los EEUU. Se recomienda una dosis de vacuna para los chicos entre 12 meses y 12 años de edad. Para los que tienen 13 años o más y no han padecido varicela, se recomiendan dos dosis administradas con un intervalo de 4 a 8 semanas. Se ha informado que el 2.8% de la población vacunada contrae de todos modos la enfermedad. Sin embargo, tales casos son generalmente leves, con un promedio de 50 lesiones, comparado con el promedio de 300 lesiones que desarrollan naturalmente los chicos no vacunados que padecen varicela.

a) Suponga que se toma una MAS de 2 chicos que no han tenido varicela y que han sido ambos vacunados. ¿Cuál es la probabilidad de que ambos niños contraigan igualmente la enfermedad?

b) La familia Pérez tiene dos hijos de 2 y 4 años de edad, que han sido vacunados ya que aún no han tenido varicela. La probabilidad calculada en a) ¿se aplica a estos dos chicos? Explique por qué sí o por qué no.

6.21. Una fábrica tiene tres máquinas que producen el 50%, 30% y 20% de la producción total respectivamente. El porcentaje de piezas defectuosas que produce cada máquina es:

Primera máquina: 4%
Segunda máquina: 6%
Tercera máquina: 2%

Si se extrae una pieza al azar de la línea de producción, ¿Cuál es la probabilidad de que sea defectuosa?

6.22. Analfabetismo. En el cuadro de la derecha aparecen datos correspondientes a 80 familias rurales del norte del país. Los datos proporcionan información sobre analfabetismo en padres e hijos.

Padre	Hijo	
	Alfabeto	Analfabeto
Alfabeto	18	7
Analfabeto	22	33

a) Los eventos "Padre analfabeto" e "Hijo analfabeto" ¿son independientes? Justifique.

b) Los eventos "Padre analfabeto" e "Hijo analfabeto" ¿son mutuamente excluyentes? Justifique.

6.23. A favor o en contra. En el Para resolver!!! 6.8 armamos el espacio muestral correspondiente a la experiencia de seleccionar al azar 3 personas y registrar si están a favor o en contra del aborto. El espacio muestral resultó:

S = {FFF, FFC, FCF, CFF, FCC, CFC, CCF, CCC}

Recordemos que el evento A lo definimos como "al menos una persona está en contra del aborto", mientras que el evento B es "exactamente dos personas están a favor del aborto".

a) ¿A y B son disjuntos? ¿Por qué sí o por qué no?
b) ¿A y B son independientes? ¿Por qué sí o por qué no?
c) Suponga que todos los resultados del espacio muestral son equiprobables.

6.24. La cantidad de números 6. Considere el experimento de tirar 2 dados. Sea X la variable correspondiente al número de 6 que salen.

S = { (1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6) }

a) ¿Cuáles son los valores posibles para la variable aleatoria X?

b) Encuentre la probabilidad de que X tome cada uno de los valores posibles.

X = x	2
P (X = x)	1/36

c) ¿Qué puede decir de las probabilidades? En particular, ¿cuánto suman?