

Indice

- **CAPÍTULO I: Conceptos básicos de probabilidad**
 - 1.1 Experimento aleatorio, espacio muestral, suceso
 - 1.2 Probabilidad
 - 1.3 Probabilidad condicional
 - 1.4 Independencia estadística
 - 1.5 Probabilidad total
 - 1.6 Regla de Bayes

- **CAPÍTULO II: Variable aleatoria unidimensional**
 - 2.1 Variable aleatoria
 - 2.2 Variables aleatorias discretas y continuas
 - 2.3 Distribución de probabilidad
 - 2.4 Función de variables aleatorias (cambio de variables)
 - 2.5 Esperanza
 - 2.6 Varianza y desvío estándar
 - 2.7 Variable aleatoria mixta
 - 2.8 Variable aleatoria condicionada o truncada
 - 2.9 Variable aleatoria mezcla

- **CAPÍTULO III: Variable aleatoria bidimensional y n-dimensional**
 - 3.1 Variable aleatoria bidimensional y n-dimensional
 - 3.2 Distribución de probabilidad conjunta
 - 3.3 Distribuciones marginales
 - 3.4 Distribuciones condicionales
 - 3.5 Independencia de variables aleatorias
 - 3.6 Esperanza condicional y regresión
 - 3.7 Esperanza, varianza, covarianza y correlación
 - 3.8 Función de variables aleatorias (cambio de variables)
 - 3.9 Distribución del máximo y el mínimo

- **CAPÍTULO IV: Proceso de Bernoulli**
 - 4.1 Experimento y proceso de Bernoulli
 - 4.2 Distribución binomial
 - 4.3 Distribución geométrica
 - 4.4 Distribución de Pascal

- **CAPÍTULO V: Proceso de Poisson**
 - 5.1 Proceso de Poisson
 - 5.2 Distribución de Poisson
 - 5.3 Distribución exponencial negativa
 - 5.4 Distribución gamma

- CAPÍTULO VI: Distribución normal y teorema central del límite
 - 6.1 Variable aleatoria normal
 - 6.2 Teorema central del límite
 - 6.3 Aproximación de binomial y Poisson por normal

- CAPÍTULO VII: Otras distribuciones particulares
 - 7.1 Distribución multinomial
 - 7.2 Distribución hipergeométrica
 - 7.3 Distribución uniforme continua
 - 7.4 Distribución ji-cuadrado.
 - 7.5 Distribución t-Student.
 - 7.6 Distribución F
 - 7.7 Distribución beta.

- CAPÍTULO VIII: Estimadores
 - 8.1 Estimadores
 - 8.2 Máxima verosimilitud
 - 8.3 Estimadores más comunes

- CAPÍTULO IX: Intervalo de confianza
 - 9.1 Intervalos de confianza
 - 9.2 Intervalo de confianza para la media de una población
 - 9.3 Intervalo de confianza para la varianza y el desvío de una población
 - 9.4 Intervalo de confianza para una proporción
 - 9.5 Intervalo de confianza para la diferencia de dos medias
 - 9.6 Intervalo de confianza para la diferencia de dos proporciones

- CAPÍTULO X: Ensayo de hipótesis
 - 10.1 Ensayos de hipótesis
 - 10.2 Ensayos de hipótesis para la media de una población.
 - 10.3 Ensayos de hipótesis para una proporción.
 - 10.4 Ensayos de hipótesis para la varianza de una población.
 - 10.5 Ensayos de hipótesis para la diferencia de dos medias.
 - 10.6 Ensayos de hipótesis para la diferencia de dos proporciones
 - 10.7 Ensayos de hipótesis para comparar dos varianzas desconocidas.
 - 10.8 Prueba de bondad de ajuste

- CAPÍTULO XI: Estimación bayesiana
 - 11.1 Estimación bayesiana
 - 11.2 Distribuciones particulares

- APÉNDICE A: Cálculo combinatorio
 - A.1 Cálculo combinatorio
 - A.2 Aplicación: Estudio de los juegos de azar

- APÉNDICE B: Otros problemas de probabilidad
 - B.1 Suma de cantidades desconocidas de probabilidades
 - B.2 Considerar varias distribuciones al mismo tiempo

- APÉNDICE C: Simulación
 - Para generar una muestra
 - Para calcular una probabilidad

- APÉNDICE D: Tablas
 - D.1 Normal estándar acumulada
 - D.2 Fractiles de la normal estándar
 - D.3 Fractiles de la t-Student
 - D.4 Fractiles de la chi-cuadrada
 - D.5 Fractiles de la F

- APÉNDICE E: Resumen de fórmulas

CAPÍTULO I

Experimento aleatorio, Espacio muestral, Suceso

Experimento Aleatorio

Definición: Es una acción o proceso que puede tener distintos resultados posibles, y cuyo resultado no se conoce hasta que no se lleva a cabo.

Ejemplos:

- tirar una moneda
- tirar un dado
- extraer una bolilla de un bolillero
- medir la cantidad de milímetros de lluvia caídos
- elegir un número al azar

Espacio muestral

Definición: Es el conjunto de resultados posibles de un experimento.

Ejemplos:

1) Si el experimento consiste en arrojar un dado y observar el número que sale, el espacio muestral es:

$$E = \{1,2,3,4,5,6\}$$

Vemos que el espacio muestral se denota con la letra E.

2) Si el experimento consiste en tomar una lapicera y medirla, el espacio muestral es:

$$E = \{ x / x \in \mathbb{R}^+ \}$$

Vemos que el espacio muestral no tiene por qué ser un conjunto finito. Como en este caso el resultado puede ser cualquier número real positivo, E tiene infinitos elementos.

3) Si el experimento consiste en tomar un libro al azar de la biblioteca y ver con qué letra empieza el título, el espacio muestral es:

$$E = \{ A, B, C, D, E, F, G, H, I, J, K, L, M, N, \tilde{N}, O, P, Q, R, S, T, U, V, W, X, Y, Z \}$$

Vemos que los resultados posibles del experimento, es decir, los elementos del espacio muestral, no tienen necesariamente por qué ser números. En este caso son letras.

4) Si el experimento consiste en tirar una moneda y ver qué sale, el espacio muestral es:

$$E = \{ \text{cara, ceca} \}$$

Aunque también podríamos haber respondido $E = \{ \text{cara, ceca, canto} \}$ si consideráramos como un resultado posible el caso en que la moneda caiga de canto

Vemos que el conjunto de resultados posibles para un experimento es subjetivo.

Generalmente adecuamos el espacio muestral a lo que consideramos posible o no posible, y a los fines del experimento. Por ejemplo, en este caso una solución posible es definir $E = \{ \text{cara, ceca} \}$ y determinar que si cae de canto, se tira nuevamente.

Esto nos lleva a la siguiente cuestión:

Distintos espacios muestrales de un mismo experimento

Como vimos en el último ejemplo, dado un experimento, no hay un único e inapelable espacio muestral asociado. De hecho el espacio muestral que definimos para un determinado experimento es arbitrario. Hay dos aspectos involucrados en dicha cuestión:

1) ¿Cuáles resultados son posibles y cuáles imposibles?

Eso es lo que ilustramos en el ejemplo anterior.

2) ¿Cómo se escriben los resultados?

Este aspecto, quizás el más trivial, se ve reflejado por ejemplo en el experimento "elegir un mes al azar", cuyo espacio muestral puede ser $E = \{ \text{enero, febrero, marzo, abril, mayo, junio, julio, agosto, septiembre, octubre, noviembre, diciembre} \}$ o bien $E = \{ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 \}$

3) ¿Qué es un resultado?

Supongamos que ya decidimos que los únicos resultados posibles al tirar una moneda son "cara" y "ceca". Ilustraremos los espacios muestrales, para mayor claridad.

- Experimento 1: "tirar una moneda y ver qué sale"

$$E = \left\{ \begin{array}{c} \text{10} \\ \text{centavos} \end{array}, \begin{array}{c} \text{Ceca} \end{array} \right\}$$

En palabras: "puede salir cara, o puede salir ceca".

Hay 2 resultados posibles.

- Experimento 2: "tirar dos monedas y ver qué sale"

$$E = \left\{ \begin{array}{cc} \text{10} & \text{10} \\ \text{centavos} & \text{centavos} \end{array}, \begin{array}{cc} \text{Ceca} & \text{Ceca} \end{array}, \begin{array}{cc} \text{10} & \text{Ceca} \\ \text{centavos} & \end{array} \right\}$$

En palabras: "pueden salir dos caras, dos cecas, o una y una".

Hay 3 resultados posibles.

- Experimento 3: "tirar una moneda de 10 centavos y una de 25 centavos y ver qué sale"

$$E = \left\{ \begin{array}{cc} \text{10} & \text{25} \\ \text{centavos} & \text{centavos} \end{array}, \begin{array}{cc} \text{10} & \text{Ceca} \\ \text{centavos} & \end{array}, \begin{array}{cc} \text{Ceca} & \text{25} \\ \text{centavos} & \text{centavos} \end{array}, \begin{array}{cc} \text{Ceca} & \text{Ceca} \end{array} \right\}$$

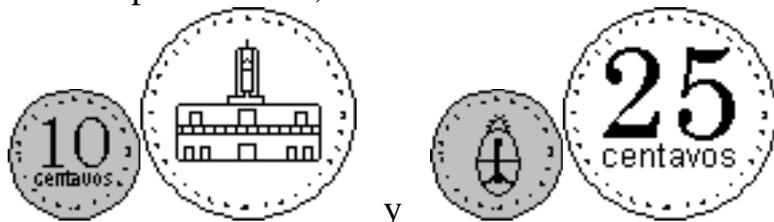
En palabras: "puede salir cara en la de 10 y cara en la de 25, cara en la de 10 y ceca en la de 25, ceca en la de 10 y cara en la de 25, ó ceca en la de 10 y ceca en la de 25".

Hay 4 resultados posibles.

¿Cómo se explica que si tanto en el experimento 2 como en el 3 arrojamos exactamente dos monedas, haya distinta cantidad de resultados posibles?

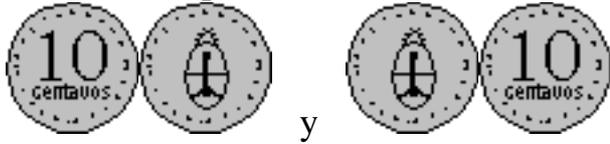
La diferencia está en que en el experimento 2, las monedas son iguales, y en el experimento 3 son distintas.

En el experimento 3, los resultados:



son, obviamente, distintos.

Pero en el experimento 2, como las monedas son iguales, los resultados:



no son **distinguibiles**, y entonces SON el mismo resultado ("una y una").

Sin embargo esto también es subjetivo, ya que esos resultados **no-distinguibiles**, pueden volverse **distinguibiles** si consideramos, por ejemplo, el orden en que se tiran las monedas, y entonces podemos tener los resultados **distinguibiles** "salió cara en la primera y ceca en la segunda" y "salió ceca en la primera y cara en la segunda".

En conclusión, al describir el espacio muestral de un experimento, es fundamental tener bien claro cuáles resultados serán **distinguibiles**, y cuáles **indistinguibiles**.

Suceso

Definición: Es un subconjunto del espacio muestral.

Ejemplos:

1) En el experimento de arrojar un dado y ver qué sale, el espacio muestral es:

$$E = \{1, 2, 3, 4, 5, 6\}$$

Cualquier subconjunto de E es un suceso, por lo tanto ejemplos de sucesos de este experimento pueden ser:

- {1}
- {6}
- {3, 4}
- {4, 5, 6}
- {1, 3, 5}
- {2, 4, 6}

También podemos expresar estos subconjuntos por comprensión:

- "que salga un número par"
- "que salga un número impar"
- "que salga un número mayor que 3"

Y no olvidemos los siguientes subconjuntos:

- {}

Dicho suceso es conocido como "suceso nulo", "suceso falso" o "**suceso imposible**". Además de la notación {} se puede usar la alternativa \emptyset .

- $\{1, 2, 3, 4, 5, 6\}$

Este subconjunto del espacio muestral es exactamente el espacio muestral (recordemos que un conjunto siempre es subconjunto de sí mismo).

Dicho suceso es conocido como "suceso verdadero", "suceso forzoso" o "**suceso cierto**".

2) En el experimento de tomar una lapicera y medir su longitud en cm.:

$$E = \{ x / x \in \mathbb{R}^+ \}$$

Ejemplos de sucesos (es decir, subconjuntos de E) pueden ser:

- $\{15\}$
- $\{14.2\}$
- $\{17.33333333\dots\}$
- $\{x \in \mathbb{R}^+ / 10 < x < 15\}$

3) Si el suceso A consiste en obtener cara al tirar una moneda, entonces podríamos definir:

- El experimento consiste en tirar una moneda y ver qué sale.
- El espacio muestral es $E = \{\text{cara, ceca}\}$
- El suceso A es $A = \{\text{cara}\}$. Vemos que $A \subset E$. Como dijimos antes, un suceso es un subconjunto del espacio muestral.

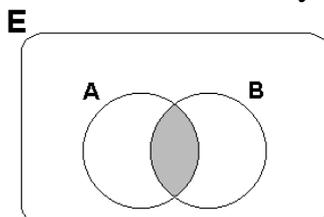
Las palabras "suceso" y "evento" se consideran sinónimas. Esto es porque habitualmente, dado un experimento, su espacio muestral E y un suceso A, si se hace el experimento, y el resultado está comprendido en el suceso A, se dice que "ocurrió" A.

Comentarios sobre los sucesos en su calidad de conjuntos

Como los sucesos son conjuntos, operar con sucesos es operar con conjuntos.

1) Intersección de sucesos

Dados A y B dos sucesos, $A \cap B$ es el suceso que ocurre cuando ocurren simultáneamente A y B. Se puede llamar "A intersección B" o bien "A y B".



Ejemplo:

Se tira un dado, y se definen los sucesos:

A: que salga menos de 4

B: que salga más de 2

Con lo cual queda:

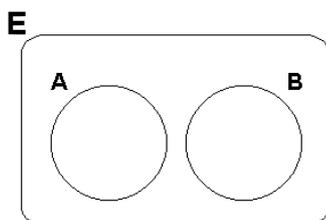
$$A = \{1, 2, 3\}$$

$$B = \{3, 4, 5, 6\}$$

$$A \cap B = \{3\}$$

2) Sucesos disjuntos o mutuamente excluyentes

Son los sucesos cuya intersección es nula. Dados los sucesos **A** y **B**, son **disjuntos** $\Leftrightarrow A \cap B = \emptyset$.



Ejemplo:

Se tira un dado, y se definen los sucesos:

A: que salga 1 ó 2

B: que salga más de 4

Con lo cual queda:

$$A = \{1, 2\}$$

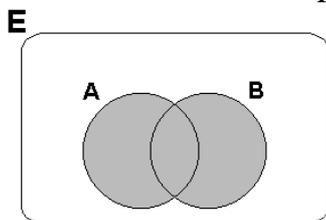
$$B = \{5, 6\}$$

$$A \cap B = \emptyset$$

Como A y B tienen intersección nula, no pueden suceder simultáneamente.

3) Unión de sucesos

Dados A y B dos sucesos, $A \cup B$ es el suceso que ocurre cuando ocurre A, B, o los dos simultáneamente. Se puede llamar "A unión B" o bien "A ó B".



Ejemplo:

Se tira un dado, y se definen los sucesos:

A: que salga menos de 4

B: que salga 2 ó 6

Con lo cual queda:

$$A = \{1, 2, 3\}$$

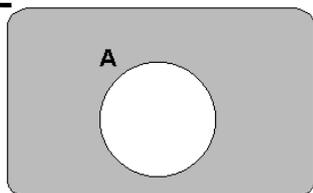
$$B = \{2, 6\}$$

$$A \cup B = \{1, 2, 3, 6\}$$

4) Complemento de los sucesos

Dado un suceso A, su "complemento" o "negado" es el suceso que ocurre si y sólo si no ocurre A (y A ocurre si y sólo si no ocurre el complemento de A). El complemento de A se escribe A^c o bien \bar{A} y se llama "complemento de A", "A negado" o bien "no A".

E



Ejemplo:

Si arrojamos un dado, y el suceso A es que salga un 4, entonces el suceso A^c es que no salga un 4 o bien que salga 1, 2, 3, 5 ó 6. Expresados como conjuntos quedan:

$$E = \{\text{sale 1, sale 2, sale 3, sale 4, sale 5, sale 6}\}$$

$$A = \{\text{sale 4}\}$$

$$A^c = \{\text{sale 1, sale 2, sale 3, sale 5, sale 6}\}$$

Observamos que:

- Así como A es un subconjunto de E, A^c también es un subconjunto de E.
- $A \cup A^c = E$, es decir, la unión de A y A^c forma E. Esto es lógico: O llueve o no llueve. No hay ninguna otra posibilidad.
- $A \cap A^c = \emptyset$. Un suceso y su complemento son disjuntos, porque no pueden ocurrir al mismo tiempo. No puede "llover" y "no llover" al mismo tiempo.

5) Partición del espacio muestral

Sea el espacio muestral E, y n sucesos A_1, \dots, A_n .

Si se cumple que:

- $A_1 \cup A_2 \cup \dots \cup A_n = E$ "la unión de los sucesos da el espacio muestral"
- $A_i \cap A_j = \emptyset \quad \forall i \neq j$ "todos los pares posibles de sucesos tienen intersección nula"

Entonces se dice que A_1, \dots, A_n forman una **partición** de E.

Como ejemplo, volvamos al experimento del dado, y definamos los siguientes sucesos:

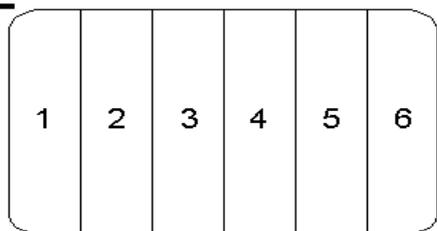
$$A_1 = \{1\}, A_2 = \{2\}, A_3 = \{3\}, A_4 = \{4\}, A_5 = \{5\}, A_6 = \{6\}.$$

Veamos que se verifica:

- $A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6 = \{1\} \cup \{2\} \cup \{3\} \cup \{4\} \cup \{5\} \cup \{6\} = \{1, 2, 3, 4, 5, 6\} = E.$
- $A_1 \cap A_2 = \emptyset, A_1 \cap A_3 = \emptyset, \dots, A_5 \cap A_6 = \emptyset.$

Entonces los A_1, \dots, A_6 que definimos forman una partición de E. Gráficamente, lo

E



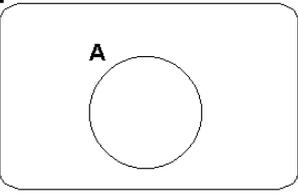
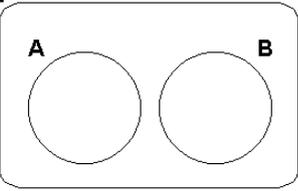
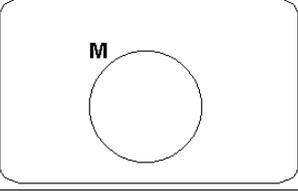
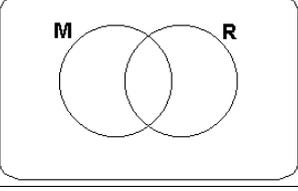
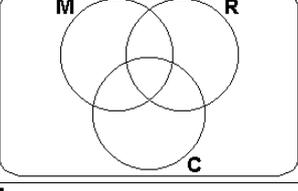
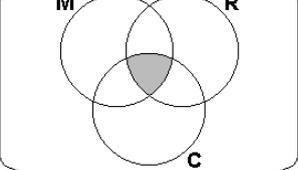
podemos ver así:

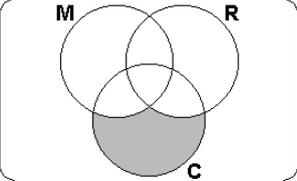
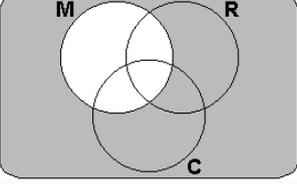
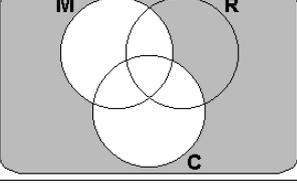
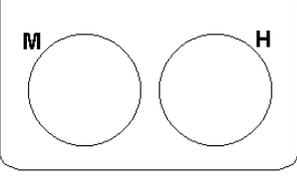
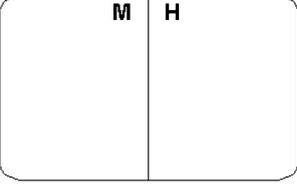
Hagamos la observación de que un suceso y su complemento siempre forman una partición del espacio muestral, puesto que como vimos antes:

- $A \cup A^c = E$
- $A \cap A^c = \emptyset$

El concepto de partición del espacio muestral nos será útil más adelante al estudiar la probabilidad total.

Como repaso, veamos algunos ejemplos gráficos:

<p>E</p> 	<p>E es el conjunto con todos los resultados posibles del experimento, y el suceso A es un subconjunto de E, porque es un conjunto de determinados resultados. Este gráfico, por ejemplo, puede estar asociado al experimento "elegir una fecha al azar" y el suceso A puede ser "el día es jueves".</p>
<p>E</p> 	<p>Al gráfico anterior le agregamos el suceso B: "la moneda es de 25 centavos". Vemos que en este ejemplo, los sucesos A y B no tienen intersección, o bien "tienen intersección nula". Es decir, son disjuntos (recordemos que A, B sucesos disjuntos $\Leftrightarrow A \cap B = \emptyset$)</p>
<p>E</p> 	<p>Cambiamos el experimento: "elegir una persona al azar". El suceso M es: "que la persona sea mujer".</p>
<p>E</p> 	<p>Al gráfico anterior le agregamos el suceso R: "que la persona sea rubia". Vemos que E queda dividido en 4 regiones: las mujeres no rubias, las mujeres rubias, los hombres rubios y los hombres no rubios.</p>
<p>E</p> 	<p>Si agregamos un tercer suceso: C = "que la persona tenga ojos claros". El espacio muestral queda dividido en 8 regiones.</p>
<p>E</p> 	<p>"Mujeres rubias de ojos claros" $M \cap R \cap C$</p>

<p>E</p> 	<p>"Hombres no-rubios de ojos claros"</p> $\overline{M} \cap \overline{R} \cap C$
<p>E</p> 	<p>"Hombres"</p> \overline{M}
<p>E</p> 	<p>"Hombres de ojos oscuros"</p> $\overline{M} \cap \overline{C}$
<p>E</p> 	<p>Si el suceso H es "la persona es hombre", entonces ese gráfico es incorrecto, a menos que sea posible no ser mujer y no ser hombre al mismo tiempo.</p> <p>Si $H = \overline{M}$, entonces $M \cup H = M \cup \overline{M}$, y como vimos antes:</p> $M \cup \overline{M} = \emptyset$
<p>E</p> 	<p>Esta es la forma correcta de hacer el gráfico anterior. Para los sucesos que forman una partición, este gráfico es correcto y el anterior no. Como vimos antes, un suceso y su complemento siempre son una partición de E. En este ejemplo:</p> $M \cup H = E ; M \cap H = \emptyset$

Problemas típicos

1) Si el suceso A es obtener un 3 al arrojar un dado, describa:

- el experimento
- el espacio muestral de dicho experimento
- el suceso A

Resolución:

- El experimento consiste en arrojar un dado.
- El espacio muestral de dicho experimento es:
 $E = \{1, 2, 3, 4, 5, 6\}$
- El suceso A es: $A = \{3\}$

Observamos que, como todo suceso, A es un subconjunto de E.

2) Si el suceso A consiste en que un número real elegido al azar entre 2 y 3 sea mayor que 2.8, describa lo mismo que se pidió en el ejercicio 1.

Resolución:

- El experimento consiste en elegir un número real al azar entre 2 y 3.
- El espacio muestral de dicho experimento es:
 $E = \{x \in \mathfrak{R} / 2 \leq x \leq 3\}$
- El suceso A es: $A = \{x \in E / x > 2.8\}$

3) Dados los experimentos descritos en 1 y 2, proponga otros sucesos para cada uno.

Resolución

- 1) Otros sucesos pueden ser: "se obtiene 6", "se obtiene menos de 4", "se obtiene más de 2", "se obtiene 3 ó 6", "no se obtiene 4", etc.
- 2) Otros sucesos pueden ser: "sale menor a 2.4", "sale entre 2.6 y 2.7", "sale exactamente 2.71", etc.

4) Describa el espacio muestral de cada uno de los siguientes experimentos aleatorios:

- a) se tira una moneda
- b) se tiran 3 monedas iguales
- c) se tiran 3 monedas distintas
- d) se tiran 2 dados iguales
- e) se tiran 2 dados distintos
- f) se eligen 2 colores primarios
- g) en una caja hay 4 bolitas negras y 1 bolita blanca. Se van sacando bolitas (sin reposición) hasta que aparezca una blanca.
- h) se coloca una pieza en un casillero al azar de un tablero de ajedrez.

Resolución:

a) $E = \{\text{cara, ceca}\}$

b) $E = \{3 \text{ caras, } 2 \text{ caras y } 1 \text{ ceca, } 1 \text{ cara y } 2 \text{ cecas, } 3 \text{ cecas}\}$

c) Si a las 2 monedas las llamamos X, Y y Z, y anotamos los resultados como $n_x n_y n_z$ donde n_i vale "a" si en la moneda i sale cara y "e" si en la moneda i sale ceca, queda:

$E = \{aaa, aae, aea, aee, eaa, eae, eea, eee\}$

Vemos que "distinguiendo" las monedas, obtenemos 8 resultados posibles, mientras que si no las distinguimos obtenemos 4 resultados posibles.

d) $E = \{2 \text{ unos, } 1 \text{ uno y } 1 \text{ dos, } 1 \text{ uno y } 1 \text{ tres, } 1 \text{ uno y } 1 \text{ cuatro, } 1 \text{ uno y } 1 \text{ cinco, } 1 \text{ uno y } 1 \text{ seis, } 2 \text{ dos, } 1 \text{ dos y } 1 \text{ tres, } 1 \text{ dos y } 1 \text{ cuatro, } 1 \text{ dos y } 1 \text{ cinco, } 1 \text{ dos y } 1 \text{ seis, } 2 \text{ tres, } 1 \text{ tres y } 1 \text{ cuatro, } 1 \text{ tres y } 1 \text{ cinco, } 1 \text{ tres y } 1 \text{ seis, } 2 \text{ cuatros, } 1 \text{ cuatro y } 1 \text{ cinco, } 1 \text{ cuatro y } 1 \text{ seis, } 2 \text{ cincos, } 1 \text{ cinco y } 1 \text{ seis, } 2 \text{ seis}\}$

e) Si a los 2 dados los llamamos X e Y, y anotamos los resultados como $n_x n_y$ donde n_x vale el número que sale en el dado X y n_y vale el número que sale en el dado Y, queda:
 $E = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, 31, 32, 33, 34, 35, 36, 41, 42, 43, 44, 45, 46, 51, 52, 53, 54, 55, 56, 61, 62, 63, 64, 65, 66\}$

Vemos otra vez que "distinguiendo" los dados, obtenemos 36 resultados posibles, mientras que si no los distinguimos obtenemos 21 resultados posibles.

f) Los colores primarios son el amarillo, el azul y el rojo.

$E = \{\text{amarillo y azul, amarillo y rojo, azul y rojo}\}$

g) Hay 2 formas de escribir el espacio muestral de este experimento.

Podemos anotar las extracciones, con lo cual los resultados posibles son:

$E = \{B, NB, NNB, NNNB, NNNNB\}$

También podemos hacer referencia al número del intento en el cual se logró sacar la blanca. Los resultados posibles son:

$E = \{1, 2, 3, 4, 5\}$

Sin duda la segunda forma es mucho más ventajosa si queremos procesar información.

h) El tablero de ajedrez tiene 8 filas (1-8) y 8 columnas (A-H). En este caso el espacio muestral puede ser las distintas "coordenadas" en las que se puede poner la ficha, con lo cual:

$E = \{A1, A2, A3, A4, A5, A6, A7, A8, B1, B2, B3, B4, B5, B6, B7, B8, C1, C2, C3, C4, C5, C6, C7, C8, D1, D2, D3, D4, D5, D6, D7, D8, E1, E2, E3, E4, E5, E6, E7, E8, F1, F2, F3, F4, F5, F6, F7, F8, G1, G2, G3, G4, G5, G6, G7, G8, H1, H2, H3, H4, H5, H6, H7, H8\}$

o bien:

$E = \{(x,y) \text{ donde } x \in \{A, B, C, E, D, F, G, H\}; y \in \{1, 2, 3, 4, 5, 6, 7, 8\}\}$

5) Un experimento consiste en lanzar un dado. Se definen 3 sucesos:

A: sale menos de 3

B: sale más de 3

C: sale 2, 4, ó 6.

Describe los siguientes sucesos:

- a) **E, A, B, C, A^c , B^c , C^c , $A \cup B$, $A \cap B$, $B \cup C$, $B \cap C$, $A \cup B^c$.**
- b) **Ocorre solamente A.**
- c) **Ocorre B, y no ocurre C.**
- d) **Ocorre alguno de los tres**
- e) **Ocurren los tres simultáneamente**
- f) **Ocorre solamente uno de los tres**
- g) **No ocurre ninguno de los tres**
- h) **Ocorre a lo sumo uno de los tres**

Resolución:

a) Nos abstraemos del hecho de que sale un número y nos quedamos directamente con los valores:

$$E = \{1, 2, 3, 4, 5, 6\}$$

$$A = \{1, 2\}$$

$$B = \{4, 5, 6\}$$

$$C = \{2, 4, 6\}$$

$$A^c = \{3, 4, 5, 6\}$$

$$B^c = \{1, 2, 3\}$$

$$C^c = \{1, 3, 5\}$$

$$A \cup B = \{1, 2, 4, 5, 6\}$$

$$A \cap B = \emptyset$$

$$B \cup C = \{2, 4, 5, 6\}$$

$$B \cap C = \{4, 6\}$$

$$A \cup B^c = \{1, 2, 3\}$$

b) "Ocorre solamente A" significa "A y noB y noC", es decir:

$$A \cap B^c \cap C^c = \{1, 2\} \cap \{1, 2, 3\} \cap \{1, 3, 5\} = \{1\}$$

Lo cual significa que si sale 1, ocurre A y solamente A.

c) "Ocorre B y no ocurre C" significa "B y noC", es decir:

$$B \cap C^c = \{4, 5, 6\} \cap \{1, 3, 5\} = \{5\}$$

Lo cual significa que si sale 5, ocurre B y no ocurre C (Y no importa si A ocurre o no).

d) "Alguno de los tres" significa "A ó B ó C", es decir:

$$A \cup B \cup C = \{1, 2, 4, 5, 6\}$$

Lo cual significa que si sale 1, 2, 4, 5 ó 6, eso garantiza que esté ocurriendo al menos uno de los tres sucesos A, B, C.

e) "Los tres simultáneamente" significa "A y B y C", es decir:

$$A \cap B \cap C = \{1, 2\} \cap \{4, 5, 6\} \cap \{2, 4, 6\} = \emptyset$$

Lo cual significa que no existe ningún número que si sale, ocurren A y B y C al mismo tiempo.

f) "Solamente uno de los 3" significa "A o bien B o bien C" (con o excluyente), lo cual es equivalente a: (A y noB y noC) o (B y noA y noC) o (C y noA y noB), es decir:

$$(A \cap B^c \cap C^c) \cup (B \cap A^c \cap C^c) \cup (C \cap A^c \cap B^c) = (\{1, 2\} \cap \{1, 2, 3\} \cap \{1, 3, 5\}) \cup (\{4, 5, 6\} \cap \{3, 4, 5, 6\} \cap \{1, 3, 5\}) \cup (\{2, 4, 6\} \cap \{3, 4, 5, 6\} \cap \{1, 2, 3\}) = \{1\} \cup \{5\} \cup \emptyset = \{1, 5\}$$

Lo cual significa que si sale 1 ó sale 5, está ocurriendo uno (y solo uno) de los 3 sucesos.

g) "Ninguno de los tres" significa noA, noB y noC, es decir:

$$(A^c \cap B^c \cap C^c) = \{3, 4, 5, 6\} \cap \{1, 2, 3\} \cap \{1, 3, 5\} = \{3\}$$

Lo cual significa que si sale 3, no está ocurriendo ninguno de los 3 sucesos.

h) "A lo sumo uno de los tres" significa "o ninguno, o uno", y eso es equivalente a "(no ocurre ninguno) ó (ocurre solamente uno). Usando los resultados obtenidos en f y g, queda:

$$\{3\} \cup \{1, 5\} = \{1, 3, 5\}$$

Lo cual significa que si sale 1, 3 ó 5, no ocurre ninguno, o a lo sumo ocurre uno.

Probabilidad

No es que hayamos estado evadiéndola, pero era necesario definir algunos conceptos y recordar ciertas cuestiones de la teoría de conjuntos antes poder responder la pregunta: **¿Qué es la probabilidad?**

- La probabilidad expresa el **grado de certeza** de que ocurrirá un determinado **suceso** al hacer un determinado **experimento aleatorio**.
- Cuanto más alta es la probabilidad de un suceso, mayor es el grado de certeza de que ocurrirá al hacer el experimento aleatorio.
- Dado un suceso A, escribimos su probabilidad como $P(A)$.

Daremos a continuación cuatro definiciones de probabilidad:

Definición informal

Informalmente, la probabilidad de un suceso es un número real entre 0 y 1.

Dicho número se puede expresar por ejemplo como 0.2, aunque también se lo puede representar como fracción ($\frac{1}{5}$), o bien como porcentaje (20%).

Si la probabilidad es 0, se **sabe** que el suceso no ocurrirá.

Si la probabilidad es 1, se **sabe** que el suceso ocurrirá.

Es decir, el 0 y el 1 son los casos **límite**.

Para valores intermedios, el suceso puede o no ocurrir. En general diremos que una probabilidad cercana a 0 es baja, y que una probabilidad cercana a 1 es alta.

Si por ejemplo la probabilidad de que mañana llueva es 0.9 significa que mañana es **altamente probable** que llueva. Si en cambio la probabilidad de que un avión se caiga es 0.00000001 significa que viajar en avión es bastante seguro.

¿Cuándo es alta una probabilidad? ¿Cuándo es baja? Eso es subjetivo. Por ejemplo si al despertarnos a la mañana el pronosticador del tiempo dice que hay 90% de probabilidades de lluvia, seguramente consideraremos que es un número alto, o por lo menos lo suficientemente alto como para tomarnos la molestia de llevar un paraguas al salir. En cambio si la probabilidad de que un avión complete un viaje sin caerse fuera ese mismo 0.9, dudo mucho que alguien quiera viajar en ese avión. Entonces cuándo una

probabilidad es o no alta o baja depende en gran medida del contexto. Es decir, **a qué esté asociada** esa probabilidad.

Ejemplos:

- 1) Si el suceso A consiste en obtener cara al tirar una moneda, entonces intuitivamente podemos decir que si la moneda no está cargada, entonces $P(A) = 1/2$.
- 2) Si el suceso A consiste en obtener un 3 al tirar un dado honesto (no cargado) entonces intuitivamente podemos decir que $P(A) = 1/6$.
- 3) Si el experimento consiste en tomar a la primera persona que veamos y preguntarle el día de la semana en que nació (supongamos que no la conocemos) entonces si el suceso A es que la persona haya nacido durante un fin de semana, diríamos intuitivamente que $P(A) = 2/7$.

Esto nos lleva a la segunda definición que daremos de probabilidad:

Definición de Laplace

En los 3 ejemplos anteriores lo que hicimos intuitivamente fue contar la cantidad de casos posibles, y luego contar la cantidad de casos contenidos en el suceso A, y responder que $P(A)$ era el cociente entre la cantidad de casos favorables a A y la cantidad de casos totales. Es decir:

$$P(A) = \frac{\text{cantidad de resultados contenidos en A}}{\text{cantidad total de resultados}}$$

Esto hace parecer que siempre que sepamos la cantidad de resultados posibles de un experimento y la cantidad de resultados englobados por el suceso A podemos calcular $P(A)$. Sin embargo, esto es falso.

Volvamos al ejemplo de las monedas:

1) ¿Cuál es la probabilidad de sacar cara al tirar una moneda no cargada?

De acuerdo al razonamiento intuitivo anterior, los resultados posibles son:

$E = \{$  ,  $\}$

Luego, si el suceso A consiste en sacar cara, constituye 1 entre 2 resultados posibles, y en consecuencia $P(A) = 1/2$.

2) ¿Cuál es la probabilidad de sacar dos caras al tirar dos monedas iguales?

Los resultados posibles son:

$$E = \left\{ \begin{array}{c} \text{10 Centavos} \quad \text{10 Centavos} \\ \text{10 Centavos} \quad \text{Cara} \\ \text{Cara} \quad \text{Cara} \end{array} \right\}$$

Entonces si A es "sacar dos caras", deberíamos decir que sacar dos caras es 1 entre 3 resultados posibles, y entonces $P(A) = 1/3$. Pero ese resultado es incorrecto, ya que intuitivamente sabemos (o deberíamos saber) que el resultado correcto es $1/4$, y que el error se debió a que tendríamos que haber usado el espacio muestral:

$$E = \left\{ \begin{array}{c} \text{10 Centavos} \quad \text{10 Centavos} \\ \text{10 Centavos} \quad \text{Cara} \\ \text{Cara} \quad \text{10 Centavos} \\ \text{Cara} \quad \text{Cara} \end{array} \right\}$$

que tiene 4 resultados posibles en vez de 3. Luego diremos correctamente que $P(A) = 1/4$.

Pero... ¿Cuál es la razón por la cual el espacio muestral que escribimos al final es apropiado y el anterior no? ¿Por qué la cantidad de resultados "correcta" es 4 y no 3, si según los que dijimos antes, ambas son formas perfectamente válidas de escribir el espacio muestral?

Y la respuesta es: porque los 4 resultados de la última expresión para E son equiprobables, mientras que los 3 de la expresión anterior no lo son.

¿Qué significa que los resultados de E sean equiprobables?

Que tienen todos la misma probabilidad.

¿Y cómo se sabe si los resultados que componen una determinada expresión de E son equiprobables?

No se sabe. Se supone.

Lamentablemente, en los problemas reales no existe una forma idónea de determinar si una determinada expresión de E está compuesta por sucesos equiprobables.

En el ejemplo de las 2 monedas, *suponemos* intuitivamente que los 4 resultados que se obtienen al diferenciar las dos monedas son equiprobables y los 3 resultados que obtienen sin distinguirlos no son equiprobables, porque el suceso "1 cara y 1 ceca" tiene 2 formas distintas de ocurrir, mientras que "2 caras" y "2 cecas" tienen solamente una forma de ocurrir cada una.

Es aceptable suponer equiprobabilidad cuando no se tiene absolutamente ningún conocimiento acerca de las probabilidades de los resultados, y eso incluye no solamente

no conocer ninguna de las probabilidades sino también no tener razones que hagan pensar que algunos resultados pueden ser más probables que otros. Eso fue lo que hicimos en el ejemplo de preguntarle a la persona el día de la semana en que nació: como no conocemos a la persona, no tenemos forma de saber qué día de la semana nació, y tampoco conocemos nada que nos pueda dar una idea de cuáles días pueden ser más probables que otros. En cambio si la pregunta fuera sobre el año de nacimiento, ya no sería tan aceptable suponer equiprobabilidad, porque no todos los años posibles tienen la misma probabilidad: por ejemplo si la persona parece ser adulta, los años recientes tienen menos probabilidad de ser el año de nacimiento de la persona que los años no-tan-recientes.

Pero entonces, ¿Cómo se pueden calcular las probabilidades cuando no se puede suponer equiprobabilidad?

Hay dos formas: una consiste en aplicar alguno de los modelos que veremos a lo largo de esta obra. La otra, tiene que ver con la tercera definición:

Definición empírica

Esta definición consiste en asociar las probabilidades de los resultados con sus frecuencias relativas luego de repetir el experimento una determinada cantidad de veces. De ahí el nombre "empírica".

Es decir,

$$P(A) \approx fr_{rel}(A) = \frac{fr_{abs}(A)}{n}$$

donde $fr_{abs}(A)$ es la cantidad de veces que ocurrió A en las n veces que se llevó a cabo el experimento.

Cuanto más grande sea n, mejor será la aproximación de $P(A)$ por $fr_{rel}(A)$.

Ejemplo:

Si se quiere tener una idea de cuál es la probabilidad de que eligiendo un alumno de la facultad al azar, éste tenga ojos claros, se puede tomar a 50 alumnos al azar y contar cuántos tienen ojos celestes. Luego si 13 de esos 50 tienen ojos claros, estimaremos que $P(A) = 13/50 = 0.26$.

Si en vez de examinar a 50 alumnos hubiéramos examinado a 200, la exactitud esperable sería mayor. Por ejemplo quizás entre los 200 alumnos habría 53 con ojos claros, y entonces $P(A) = 0.265$.

Y si hubiera infinitos alumnos, y tomáramos muestras cada vez mayores, nos acercaríamos asintóticamente al resultado real, que podría ser, por ejemplo, 0.263.

Definición axiomática

Las tres definiciones que dimos hasta ahora cumplen con esta cuarta y última definición. La definición axiomática consta de los siguientes tres axiomas:

- **Axioma 1: $P(A) \geq 0$**

"La probabilidad no puede ser negativa"

- **Axioma 2: $P(E) = 1$**

"La probabilidad del espacio muestral es uno"

- **Axioma 3: $A \cap B = \emptyset \Leftrightarrow P(A \cup B) = P(A) + P(B)$**

"Dos sucesos son disjuntos si y sólo si la probabilidad de su unión es la suma de sus probabilidades".

De los tres axiomas, se deducen casi inmediatamente cinco consecuencias:

- **Consecuencia 1: $P(A) \leq 1$**

"La probabilidad tampoco puede ser mayor que uno"

Porque como $A \subset E$, si $P(A) > 1$ entonces necesariamente $P(E) > 1$, lo cual va en contra del segundo axioma.

- **Consecuencia 2: $P(A) + P(\bar{A}) = 1$**

"Las probabilidades de dos sucesos complementarios suman uno"

$P(E) = P(A \cup \bar{A})$ porque como vimos antes $A \cup \bar{A} = E$

$P(A \cup \bar{A}) = P(A) + P(\bar{A})$ por el tercer axioma, porque A y \bar{A} son disjuntos.

y como $P(E) = 1$, $P(A) + P(\bar{A}) = 1$

Esto es muy útil porque a menudo es más fácil calcular $P(\bar{A})$ que $P(A)$, y entonces $P(A)$ se obtiene de $P(A) = 1 - P(\bar{A})$

- **Consecuencia 3: $P(\emptyset) = 0$**

"La probabilidad de un suceso imposible es cero"

Intuitivamente, si un suceso es el conjunto vacío, es porque no contiene ningún resultado, y entonces nunca podría suceder (de ahí el nombre "imposible").

Como $\emptyset \cap \emptyset = \emptyset$, entonces por el tercer axioma:

$$P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset)$$

$$P(\emptyset) = P(\emptyset) + P(\emptyset)$$

$$P(\emptyset) - P(\emptyset) = P(\emptyset)$$

$$P(\emptyset) = 0$$

- **Consecuencia 4: $A \subset B \Rightarrow P(A) \leq P(B)$**

"Si un suceso está incluido en otro, su probabilidad es a lo sumo la de éste"

Partimos B en $A \cap B$ y $\bar{A} \cap B$ y aplicamos el tercer axioma:

$$P((A \cap B) \cup (\bar{A} \cap B)) = P(A \cap B) + P(\bar{A} \cap B)$$

$$P(B) = P(A \cap B) + P(B \cap \bar{A})$$

Partimos A en $A \cap B$ y $A \cap \bar{B}$ y aplicamos el tercer axioma:

$$P((A \cap B) \cup (A \cap \bar{B})) = P(A \cap B) + P(A \cap \bar{B})$$

$$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

Pero como $A \subset B$, entonces $A \cap \bar{B} = \emptyset$, con lo cual $P(A \cap \bar{B}) = 0$, y entonces queda:

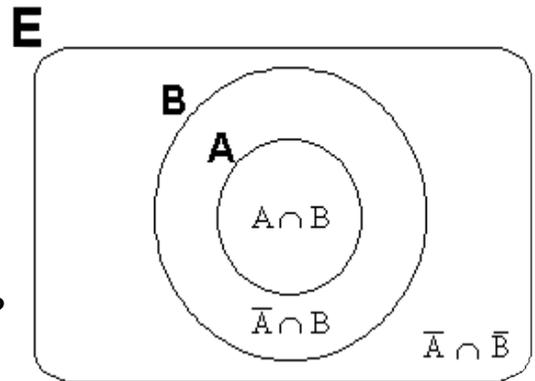
$$P(A) = P(A \cap B)$$

Y como, según calculamos antes, $P(B) = P(A \cap B) + P(B \cap \bar{A})$, queda:

$$P(A) = P(B) - P(B \cap \bar{A})$$

Y como $P(B \cap \bar{A}) \geq 0$, llegamos lo que queríamos demostrar.

Observemos que en el caso particular de que A no solamente esté incluido en B sino que sea igual a B (la igualdad de conjuntos es un caso particular de inclusión) entonces queda $P(B \cap \bar{A}) = 0$ y consecuentemente $P(A) = P(B)$.



• **Consecuencia 5: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$**

La probabilidad de la unión de dos sucesos es la suma de sus probabilidades menos la probabilidad de la intersección.

Tomemos la siguiente partición de E: $\{C_1, C_2, C_3, C_4\}$ donde $C_1 = A \cap B$, $C_2 = A \cap \bar{B}$, $C_3 = \bar{A} \cap B$, $C_4 = \bar{A} \cap \bar{B}$

Luego:

$$A = C_1 \cup C_2 \text{ por propiedades de conjuntos}$$

$$B = C_1 \cup C_3 \text{ por propiedades de conjuntos}$$

$$P(A) = P(C_1) + P(C_2) \text{ por el tercer axioma}$$

$$P(B) = P(C_1) + P(C_3) \text{ por el tercer axioma}$$

$$A \cup B = C_1 \cup C_2 \cup C_3 \text{ por propiedades de conjuntos}$$

$$P(A \cup B) = P(C_1) + P(C_2) + P(C_3) \text{ por el tercer axioma dos veces}$$

$$A \cap B = C_1 \text{ por propiedades de conjuntos}$$

$$P(A \cap B) = P(C_1) \text{ porque si } X = Y \text{ entonces } P(X) = P(Y)$$

Juntando todo queda que:

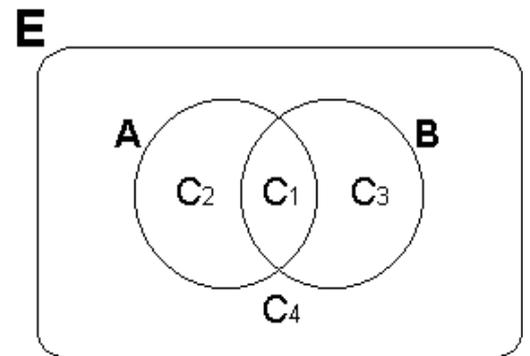
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

es equivalente a:

$$P(C_1) + P(C_2) + P(C_3) = P(C_1) + P(C_2) + P(C_1) + P(C_3) - P(C_1)$$

Simplificando del lado derecho:

$$P(C_1) + P(C_2) + P(C_3) = P(C_1) + P(C_2) + P(C_3)$$



Con lo cual la tercera consecuencia es válida.

Explicación intuitiva: Al construir $A \cup B$ "sumando" A y B estamos "contando" dos veces la intersección; por eso hay que restarla. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Cuando son disjuntos (el caso contemplado por el tercer axioma) la intersección es \emptyset , por eso en la expresión del axioma no hace falta que aparezca restando.

Generalización de la quinta consecuencia:

• **Para 3 sucesos:**

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

"La probabilidad de la unión de tres sucesos es:

las probabilidades individuales

menos las probabilidades de las intersecciones tomadas de a 2

más la probabilidad de la intersección tomada de a 3"

Análogamente:

• Para 4 sucesos:

"La probabilidad de la unión de cuatro sucesos es:

1) Las probabilidades individuales (sumando)

2) menos las probabilidades de las intersecciones tomadas de a 2

3) más las probabilidades de las intersecciones tomadas de a 3

4) menos la probabilidad de la intersección tomada de a 4"

Y así sucesivamente, alternando el signo se puede obtener la forma de calcular la probabilidad de la unión de cualquier número de sucesos.

Problemas típicos

1) **Se tiran dos dados no cargados. Indique la probabilidad de que:**

a) **Salgan dos 3**

b) **Salgan dos 4**

c) **No salga ningún 5**

d) **Salga algún 5**

e) **No salga ningún 5 ni ningún 6**

f) **Salgan solamente números pares**

Resolución

El espacio muestral es el siguiente:

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

Usamos este espacio muestral porque suponemos que sus elementos son equiprobables. Si hubiéramos considerado los dos dados no-distinguibles, entonces el suceso (1,2) tendría 2 formas posibles de ocurrir, y como vimos en el ejemplo de las monedas eso nos condujo a un espacio muestral no-equiprobable.

Queremos que el espacio muestral sea equiprobable para poder aplicar la definición de Laplace.

Hay 36 formas posibles de tirar los dos dados. Luego contando los resultados incluidos en cada suceso cuya probabilidad se pide, obtenemos:

- a) $1/36$
- b) $1/36$
- c) $25/36$
- d) "salga algún 5" quiere decir "al menos un 5", es decir, 1 ó 2 cincos. En otras palabras, es el complemento del suceso a anterior. Su probabilidad es $11/36$
- e) $16/36$
- f) $9/36$

2) En una determinada población, el 60% de las personas son mujeres, el 35% de la gente tiene ojos claros y el 25% de la gente es rubia. El 20% de la población son mujeres de ojos claros. El 10% de la población son mujeres rubias. El 15% de la población son personas rubias y de ojos claros. El 5% de la población son mujeres rubias de ojos claros. Calcule las probabilidades de que al elegir una persona al azar, esta:

- a) sea mujer, sea rubia o tenga ojos claros (es decir, que tenga por lo menos una de esas 3 características).
- b) tenga ojos oscuros
- c) sea un hombre no rubio y de ojos oscuros
- d) tenga cabello rubio o no tenga cabello rubio (alguna de las dos cosas).
- e) tenga ojos claros y ojos oscuros (las dos cosas simultáneamente).
- f) La probabilidad de encontrar a una mujer rubia, ¿es menor, igual, o mayor, que la de encontrar a una mujer rubia de ojos claros?

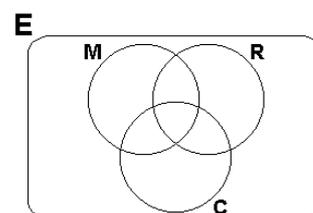
Resolución

Definiremos los sucesos:

- M: la persona es mujer
- R: la persona es rubia
- C: la persona tiene ojos claros

Entonces los datos son:

$$\begin{array}{lll}
 P(M) = 0.6 & P(C) = 0.35 & P(R) = 0.25 \\
 P(M \cap C) = 0.2 & P(M \cap R) = 0.1 & P(R \cap C) = 0.15 \\
 P(M \cap C \cap R) = 0.05 & &
 \end{array}$$



Vamos a resolver el ejercicio de 3 formas distintas.

• **Forma 1: Aplicando los axiomas de la probabilidad y sus consecuencias para hallar las probabilidades pedidas.**

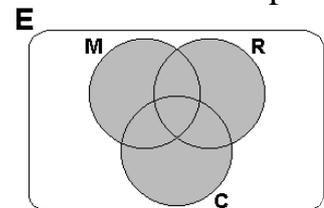
a) Nos piden $P(M \cup C \cup R)$. Por la generalización de la quinta sucesos, sabemos que:

$$P(M \cup C \cup R) = P(M) + P(C) + P(R) - P(M \cap C) - P(M \cap R) - P(C \cap R) + P(M \cap C \cap R)$$

Y en este caso, todos los sumandos del lado derecho de la igualdad son dato. Entonces obtenemos:

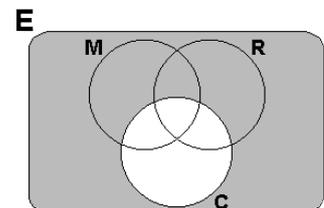
$$P(M \cup C \cup R) = 0.6 + 0.35 + 0.25 - 0.2 - 0.1 - 0.15 + 0.05 = \mathbf{0.8}$$

consecuencia para 3



b) El suceso "tener ojos oscuros" es la negación del suceso "tener ojos claros". Es decir, es el complemento de C. La segunda consecuencia nos dice que $P(A) + P(\bar{A}) = 1$, con lo cual:

$$P(\bar{C}) = 1 - P(C) = 1 - 0.35 = \mathbf{0.65}$$



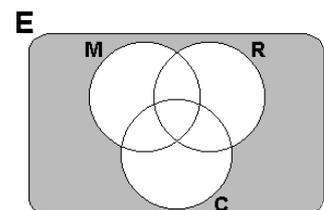
c) Aquí el razonamiento es similar al del punto anterior. Si la persona elegida es hombre, no-rubio, y de ojos oscuros, no tiene ninguna de las 3 características M, C y R, y salió el complemento del conjunto $M \cup C \cup R$ (lo de afuera de los tres globos del diagrama de Venn).

La segunda consecuencia dice que $P(A) + P(\bar{A}) = 1$, con lo cual si llamamos:

$$A = M \cup C \cup R$$

entonces lo que estamos buscando es $P(\bar{A})$, y como conocemos $P(A)$, hacemos:

$$P(\bar{A}) = 1 - P(A) = 1 - 0.8 = \mathbf{0.2}$$



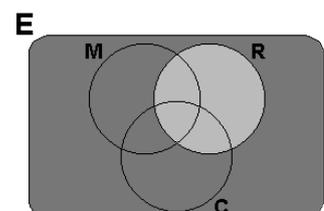
d) Estamos buscando $P(R \cup \bar{R})$. Como los sucesos complementarios son disjuntos (porque necesariamente $A \cap \bar{A} = \emptyset$), por el tercer axioma:

$$P(R \cup \bar{R}) = P(R) + P(\bar{R}).$$

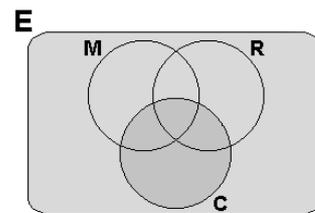
Luego por la segunda consecuencia:

$$P(R) + P(\bar{R}) = \mathbf{1}$$

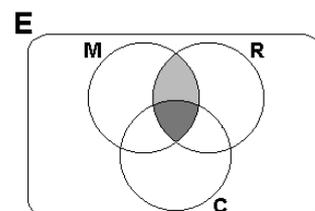
Este resultado era evidente, porque sólo se puede ser rubio o no-rubio. Sólo puede llover o no-llover. Por lo tanto la probabilidad de que suceda alguna de las dos cosas es necesariamente 1, porque siempre sucede alguna de las dos cosas.



e) Nos piden $P(C \cap \bar{C})$. C y su complemento no pueden ocurrir al mismo tiempo, porque una persona no puede tener ojos claros y ojos no-claros simultaneamente (supongamos que las personas tienen los dos ojos del mismo color). Entonces como las dos cosas no pueden ocurrir al mismo tiempo, la probabilidad de su intersección es necesariamente **cero**.



f) Las mujeres rubias pueden tener ojos claros u ojos oscuros. Siempre que una mujer sea rubia y de ojos claros, será necesariamente mujer rubia, pero no al revés, porque el hecho de que una mujer sea rubia no garantiza que además tenga ojos claros. Entonces la probabilidad de encontrar una mujer rubia que además tenga ojos claros es menor que la probabilidad de simplemente encontrar a una mujer rubia.



Si lo queremos pensar por la cuenta consecuencia:

$$(M \cap R \cap C) \subset (M \cap R) \Rightarrow P(M \cap R \cap C) < P(M \cap R)$$

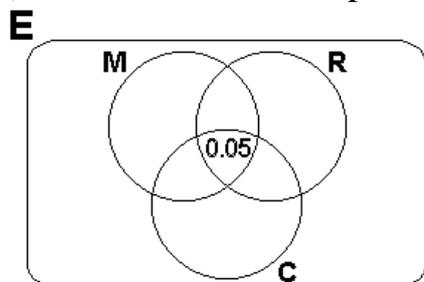
(usamos $<$ y no \leq porque \leq es para el caso particular en el cual un conjunto está incluido en otro porque ambos conjuntos son iguales (recordemos que $A = B \Rightarrow A \subset B$ y $B \subset A$))

• **Forma 2: Aplicando los axiomas de la probabilidad y sus consecuencias para hallar todas las probabilidades.**

Siendo los datos:

$$\begin{aligned} P(M) &= 0.6 & P(C) &= 0.35 & P(R) &= 0.25 \\ P(M \cap C) &= 0.2 & P(M \cap R) &= 0.1 & P(R \cap C) &= 0.15 \\ P(M \cap C \cap R) &= 0.05 \end{aligned}$$

1) En la intersección triple tenemos 0.05



2) $(M \cap C)$ es la unión de los sucesos disjuntos:

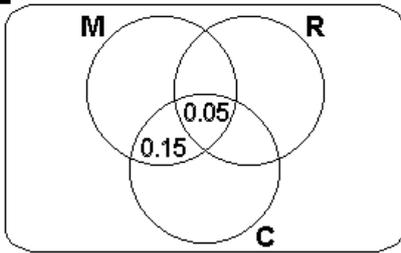
$$(M \cap C \cap R) \text{ y } (M \cap C \cap \bar{R}).$$

Luego:

$$P(M \cap C \cap R) + P(M \cap C \cap \bar{R}) = P(M \cap C)$$

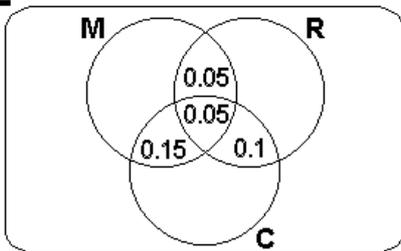
$$\Rightarrow P(M \cap C \cap \bar{R}) = P(M \cap C) - P(M \cap C \cap R) = 0.2 - 0.05 = 0.15$$

E



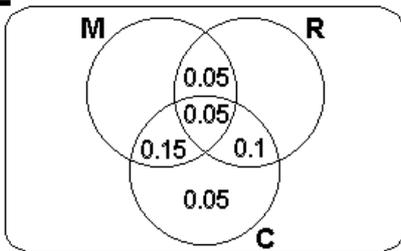
3) Análogamente aplicamos lo mismo para $(M \cap R)$ y para $(R \cap C)$. Es decir, sabemos que la probabilidad del "óvalo" $(M \cap R)$ debe dar en total 0.1, y que la probabilidad del "óvalo" $(R \cap C)$ debe dar en total 0.15.

E



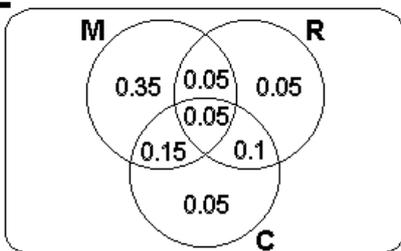
4) Sabemos que en total $P(C)$ tiene que dar 0.35, por lo cual $P(\bar{M} \cap \bar{R} \cap C)$ debe dar 0.05.

E

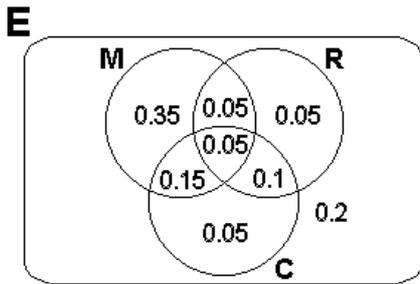


5) Análogamente hacemos lo mismo para M y para R.

E



6) Como sabemos que $P(E)$ debe dar en total 1, la probabilidad de la región que se encuentra afuera de los 3 conjuntos debe ser 0.2.

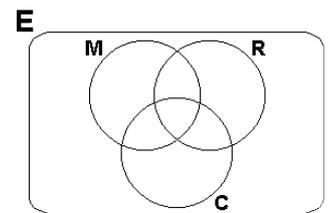


Luego las respuestas a las preguntas son inmediatas.

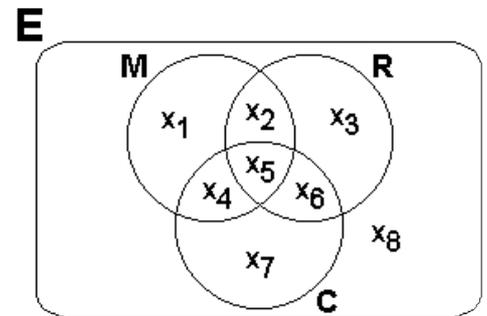
• **Forma 3: Planteando un sistema y resolviéndolo**

La tercera forma nos permite un mayor grado de automatización (que nos sería útil por ejemplo si fuéramos a desarrollar algún tipo de software que resolviera estas cuestiones).

Tomando los tres sucesos, el espacio muestral nos quedó dividido en $2^3 = 8$ regiones (el 2 porque al hacer el experimento puede pasar que ocurra o no ocurra (2 posibilidades) ese suceso, y el 3 porque eso lo aplicamos a cada uno de los 3 sucesos que estamos considerando). Tenemos entonces 8 incógnitas.



Comenzamos por ponerle nombre a cada una de las regiones. Si llamamos x_i a $P(\text{región } i)$, entonces por ejemplo nos podría quedar como vemos en el gráfico. Luego escribimos ecuaciones a partir de los datos que tenemos:



Dato	Ecuación
$P(M) = 0.6$	$x_1 + x_2 + x_4 + x_5 = 0.6$
$P(C) = 0.35$	$x_4 + x_5 + x_6 + x_7 = 0.35$
$P(R) = 0.25$	$x_2 + x_3 + x_5 + x_6 = 0.25$
$P(M \cap C) = 0.2$	$x_4 + x_5 = 0.2$
$P(M \cap R) = 0.1$	$x_2 + x_5 = 0.1$
$P(R \cap C) = 0.15$	$x_5 + x_6 = 0.15$
$P(M \cap C \cap R) = 0.05$	$x_5 = 0.05$

Podría parecer que tenemos solamente 7 ecuaciones para las 8 incógnitas, pero también sabemos que la probabilidad del espacio muestral es 1, es decir:

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 = 1$$

El sistema ampliado queda:

$$\left(\begin{array}{cccccccc|c} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0.6 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0.35 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0.25 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0.2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0.1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0.15 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0.05 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right)$$

De donde por cualquier método, por ejemplo el de Gauss, obtenemos:

$$\begin{array}{llll} x_1=0.35 & x_2=0.05 & x_3=0.05 & x_4=0.15 \\ x_5=0.05 & x_6=0.1 & x_7=0.05 & x_8=0.2 \end{array}$$

Con lo cual ya tenemos todo resuelto y estamos en condiciones de responder sobre las probabilidades de cualquiera de los 8 casos o uniones de ellos.

Para hallar las respuestas podemos sumar todas las probabilidades x_i de las regiones que cumplan con la condición. Si las regiones que cumplen con la condición son muchas, podemos hacer $1 -$ [las probabilidades de las regiones que NO cumplen con la condición].

Luego:

- a) $1 - x_8 = 0.8$
- b) $x_1 + x_2 + x_3 + x_8 = 0.65$
- c) $x_8 = 0.2$
- d) $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 = 1$
- e) 0
- f) mujer rubia: $x_2 + x_5 = 0.1$
mujer rubia de ojos claros: $x_5 = 0.05$
 $0.1 > 0.01$

Probabilidad condicional

Supongamos que estamos estudiando el rendimiento de los alumnos de la materia Probabilidad y Estadística en un determinado examen.

De un relevamiento surge que:

- el 80% de los alumnos estudió para el examen
- el 75% de los alumnos aprobó el examen
- el 15% de los alumnos no estudió para el examen y no lo aprobó.

Si definimos el experimento de tomar un alumno al azar, y llamamos A al suceso "el alumno tomado aprobó el examen" y B al suceso "el alumno tomado estudió para el examen", entonces tenemos que:

$$P(A) = 0.75$$

$$P(B) = 0.8$$

$$P(\bar{A} \cap \bar{B}) = 0.15$$

Con estos datos y considerando que $P(E) = 1$, ya podemos hacer el diagrama de Venn correspondiente y conocer las probabilidades de todas las regiones.

Por ejemplo, si quisiéramos evaluar el nivel de los profesores y las clases, nos puede interesar responder la pregunta: ¿Cuál es la probabilidad de que un alumno que haya estudiado haya aprobado el examen?

Intuitivamente podemos darnos cuenta de que, al menos bajo ciertas circunstancias, el procedimiento para encontrar la respuesta podría ser fijarnos, de entre los alumnos que estudiaron, cuántos aprobaron.

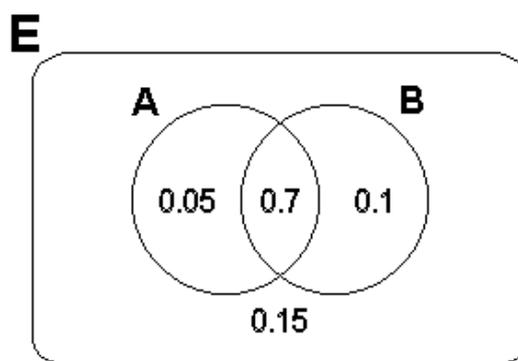
Los alumnos que estudiaron fueron el 80%.

Ese 80% está formado un 70% que aprobaron y un 10% que no aprobaron.

Entonces podemos decir que de cada 80 alumnos que estudiaron, 70 aprobaron.

Visto de otra forma, si estamos parados en B, la probabilidad de estar al mismo tiempo también parados en A es $70/80 = 0.875$.

La cuenta que hicimos intuitivamente fue calcular la proporción entre la cantidad de alumnos que [estudió y aprobó], sobre el total de alumnos que estudiaron.



Entonces, respondiendo a la pregunta, la probabilidad de que un alumno que estudie apruebe, es decir, la probabilidad de que ocurra A dado que ocurrió B, se escribe $P(A/B)$ y vale:

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

Dicha expresión constituye la definición de probabilidad condicional, y vale para todo par de sucesos A, B contenidos en el mismo espacio muestral.

$P(A/B)$ se lee "probabilidad condicional de A dado B", o bien "probabilidad de A dado B" o bien "probabilidad de que ocurra A sabiendo que ocurrió B".

Más abajo se mostrará conceptualmente cómo se llega a esa expresión.

En este ejemplo quedan definidas las siguientes probabilidades:

- $P(A)$ probabilidad de que un alumno cualquiera apruebe
- $P(B)$ probabilidad de que un alumno cualquiera estudie
- $P(A/B)$ probabilidad de que un alumno que estudió apruebe
- $P(B/A)$ probabilidad de que un alumno que aprobó haya estudiado

Y también:

- $P(A/\bar{B})$ probabilidad de que un alumno que no estudió apruebe
- $P(B/\bar{A})$ probabilidad de que un alumno que no aprobó haya estudiado
- $P(\bar{A}/B)$ probabilidad de que un alumno que estudió no apruebe
- $P(\bar{B}/A)$ probabilidad de que un alumno que aprobó no haya estudiado

A modo ilustrativo, calcularemos algunas:

La probabilidad de que un alumno que aprobó haya estudiado es la probabilidad de que ocurra B(estudió) sabiendo que ocurrió A(aprobó), es decir:

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{0.7}{0.75} = 0.933$$

Notemos que no es lo mismo la probabilidad de que un alumno que estudió apruebe ($P(A/B)$) que la probabilidad de que un alumno que aprobó haya estudiado ($P(B/A)$).

La probabilidad de que un alumno apruebe sin estudiar es la probabilidad de que apruebe dado que no estudió, es decir, la probabilidad de que ocurra A sabiendo que no ocurrió B, o sea:

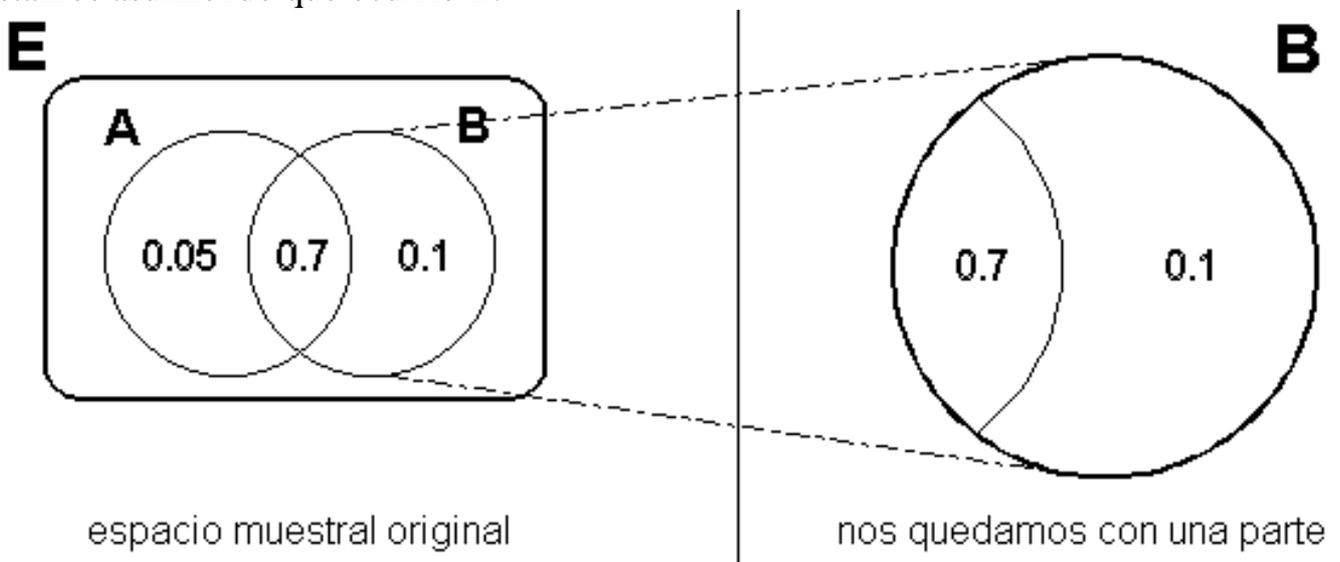
$$P(A/\bar{B}) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{0.05}{0.2} = 0.25$$

¿Cómo explicamos desde los conceptos vistos hasta ahora la expresión hallada para la probabilidad condicional?

Como vimos antes, la probabilidad de que ocurra A dado que ocurrió B es la probabilidad de que ocurran A y B simultáneamente dividida la probabilidad de que ocurra B, es decir, intuitivamente, la probabilidad de "estar parados en A, sabiendo que estamos parados en B".

Lo que sucede es que el hecho de "estar parados en B" implica que estamos asumiendo que B es cierto. Es decir, estamos calculando probabilidades a condición de que B ocurra. Eso no se diferencia en nada de considerar, al menos por un momento, que B es nuestro nuevo espacio muestral, y que $P(A/B)$ no es otra cosa que $P(A)$ dentro de ese nuevo espacio muestral.

Es decir, $P(A/B)$ es en realidad la probabilidad de que ocurra A en un espacio en el que estamos asumiendo que ocurrió B.



Pero el B con el que nos quedamos todavía no está listo para ser un espacio muestral, porque sus probabilidades no suman 1.

Justamente, las probabilidades que tienen en ese gráfico no son correctas porque estaban referidas al espacio muestral E. Hay que adaptarlas respetando dos cosas:

- Ahora deberán sumar 1.
- No se debe alterar la proporción relativa que tienen las probabilidades dentro de B.

La forma de cumplir con esas dos cuestiones es multiplicar (o dividir) todas las probabilidades que están en B por el mismo factor.

¿Cuál es ese factor? Comencemos por notar que las probabilidades contenidas en B suman $P(B)$. Entonces dividiendo todas las probabilidades por $P(B)$, la suma tiene que dar 1.

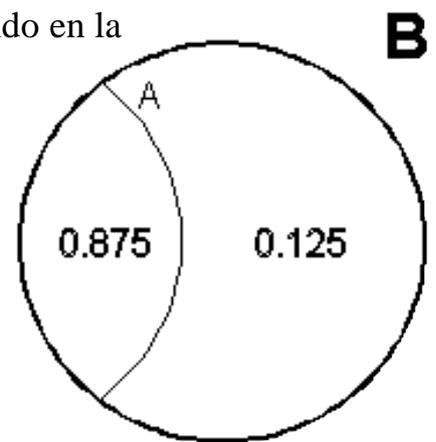
Y al estar dividiendo todas las probabilidades por el mismo número, la proporción se mantiene. Ahora ya sabemos por qué aparece el $P(B)$ dividiendo en la definición de probabilidad condicional.

En el ejemplo, $P(B) = 0.8$

Entonces el 0.7 se convierte en $0.7 / 0.8 = 0.875$

Y el 0.1 se convierte en $0.1 / 0.8 = 0.125$

Con lo cual ya tenemos todo lo que necesitamos para describir nuestro nuevo espacio muestral B.



el nuevo espacio muestral

Para aún mayor claridad, podemos cambiarle el estilo a este diagrama de Venn, de modo de hacerlo más parecido al estilo al que estamos acostumbrados:

B Vemos que hicimos para el espacio muestral B el mismo tipo de diagrama que solemos hacer para el espacio muestral E. Esto es para mostrar que podemos pararnos en un nuevo espacio muestral (que puede ser un subconjunto del espacio muestral original) y obtener un espacio muestral tan válido como el original, con la diferencia de que las probabilidades que aparecen en el nuevo espacio muestral están referidas al nuevo espacio muestral y no al original. Es por eso que el 0.875 que aparece es $P(A)$ pero **referida al espacio muestral B**, es decir, $P(A/B)$.

Si se sobreentiende que nos estamos refiriendo al espacio muestral B, entonces no hace falta escribir $P(A/B)$ y podemos escribir simplemente $P(A)$.

De hecho si lo pensamos, cuando trabajamos en el espacio muestral E, las probabilidades están referidas al espacio muestral E, pero como normalmente se sobreentiende que las probabilidades están referidas al espacio muestral E, no hace falta escribir $P(C/E)$ y escribimos directamente $P(C)$.

Una notación que se suele utilizar es colocarle como subíndice al operador P el espacio muestral al cual se refiere la probabilidad. Entonces $P(A/B)$ se puede escribir también $P_B(A)$ lo cual se lee "probabilidad de A referida al espacio muestral B" o bien exactamente igual que antes "probabilidad de A dado B".

Otra cuestión que podemos notar es que hasta ahora nunca nos habían aparecido probabilidades multiplicando o dividiendo, sino siempre sumando o restando. Las

probabilidades multiplicando o dividiendo son características de los cambios de espacio muestral, tema que hasta ahora no habíamos explorado.

Intersección de sucesos y multiplicación de probabilidades

De la definición de probabilidad condicional obtenemos en forma inmediata que:

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

Esto nos da por fin una forma de calcular probabilidades de intersecciones para los casos en que no conocemos la probabilidad de la unión y entonces no podemos usar:

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

Si pensamos $P(A \cap B)$ como $P(B \cap A)$, con la probabilidad condicional obtenemos:

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

¿Qué sucede con la intersección de 3 sucesos?

La probabilidad de la intersección es $P(A \cap B \cap C)$.

Asociando A y B, y usando probabilidad condicional, hacemos:

$$P(A \cap B \cap C) = P(C \cap (A \cap B)) = P\left(\frac{C}{(A \cap B)}\right) P(A \cap B)$$

Si ahora aplicamos que $P(A \cap B) = P(B/A) P(A)$ nos queda el siguiente resultado:

$$P(A \cap B \cap C) = P(A) P\left(\frac{B}{A}\right) P\left(\frac{C}{A \cap B}\right)$$

Para n sucesos, podemos generalizar este resultado. Si llamamos A_1, A_2, \dots, A_n a los n sucesos, nos queda:

$$P(\text{intersect } A_i) = \prod_{i=1}^n P\left(A_i / \text{intersect } A_j\right)_{j=1}^{i-1}$$

Ejemplo

El 95% de los gatos de 3 colores son hembras. El 40% de los gatos son hembras. Al tomar un gato al azar, ¿cuál es la probabilidad de que sea una hembra de 3 colores?

Si el suceso A es que el gato elegido sea de 3 colores y el suceso B es que sea hembra, estamos buscando $P(A \cap B)$. Nos dieron de dato:

$$P(A/B) = 0.95$$

$$P(B) = 0.4$$

Usando probabilidad condicional calculamos:

$$P(A \cap B) = P(A/B) \cdot P(B) = 0.95 \cdot 0.4 = 0.38$$

Ejemplo

Se tienen en una caja 3 bolitas negras y 3 bolitas blancas. ¿Cuál es la probabilidad de sacar 2 bolitas y que resulten ser blancas?

Analicemos:

Como originalmente hay 3 bolitas negras y 3 blancas, la probabilidad de sacar una bolita blanca es 0.5. Sacamos una bolita y la dejamos afuera.

Supongamos que la bolita que sacamos resultó ser blanca. ¿Cuál es ahora la probabilidad de sacar una bolita blanca? Intuitivamente (por ahora) responderemos que $\frac{2}{5}$, porque quedan 2 bolitas blancas en las 5 que hay.

Ahora le pondremos nombre a estos sucesos:

A: que la primera bolita sacada sea blanca

B: que la segunda bolita sacada sea blanca

Evidentemente lo que estamos buscando es $P(A \cap B)$

Vimos que $P(A \cap B) = P(A/B) \cdot P(B) = P(B/A) \cdot P(A)$

Y según lo que analizamos recién, conocemos $P(A) = 0.5$, y también conocemos $P(B/A)$, porque sabemos cuál es la probabilidad de que la segunda bolita sea blanca sabiendo que la primera lo fue. Habíamos determinado que era $\frac{2}{5}$. Entonces calculamos $P(A \cap B)$:

$$P(A \cap B) = P(A) \cdot P(B/A) = \frac{2}{5} \cdot 0.5 = \frac{1}{5}$$

Con lo cual podemos responder a la pregunta: la probabilidad de sacar 2 bolitas y que ambas sean blancas, es $\frac{1}{5}$.

Antes comentamos que cuando aparecían probabilidades multiplicando eso indicaba cambios de espacios muestrales. El $P(B/A)$ que usamos es la probabilidad de que ocurra B referida al espacio muestral A. Es decir, luego de que sacamos una bolita blanca, cuando llega el momento de sacar la segunda bolita el espacio muestral **ya no es el mismo** que era antes de sacar la primera (porque la composición de las bolitas en la caja ya no es la misma).

Ahora pensemos en un caso más complejo: ¿cuál es la probabilidad de sacar 3 bolitas, de modo tal que las dos primeras sean blancas, y la tercera sea negra?

Definimos un nuevo suceso:

C: que la tercera bolita sacada sea negra

Y entonces lo que estamos buscando es $P(A \cap B \cap C)$. Aplicando lo estudiado antes,

$$P(A \cap B \cap C) = P(A) P\left(\frac{B}{A}\right) P\left(\frac{C}{A \cap B}\right)$$

$P(A)$ es la probabilidad de que la primera bolita sea blanca, o sea $\frac{3}{6}$

$P(B/A)$ es la probabilidad de que la segunda bolita sea blanca, dado que la primera fue blanca. Como vimos antes, luego de sacar una bolita blanca queda 3 negras y 2 blancas, con lo cual $P(B/A) = \frac{2}{5}$.

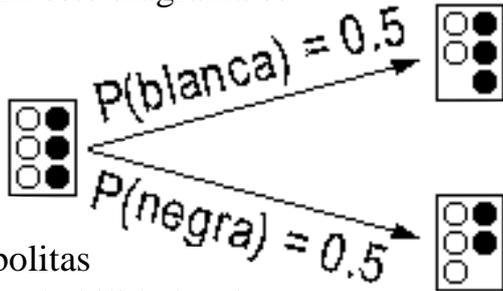
$P(C / (A \cap B))$ es la probabilidad de que la tercera bolita sea negra, dado que de la caja original se sacaron dos blancas. Al momento de sacar la tercera bolita, quedan 3 negras y una blanca, con lo cual $P(C / (A \cap B)) = 3/4$.

Luego la probabilidad buscada es:

$$P(A \cap B \cap C) = \frac{3}{6} \cdot \frac{2}{5} \cdot \frac{3}{4} = 0.15$$

Ahora veremos un diagrama que nos podrá ser de utilidad en estos casos:

En este diagrama se



bolitas

probabilidades de que

así sucesivamente. Esta lógica se puede seguir aplicando recursivamente mientras sigan quedando bolitas en la caja.

Si hiciéramos el diagrama de árbol para las primeras 3 bolitas que se extraen, el diagrama quedaría así:

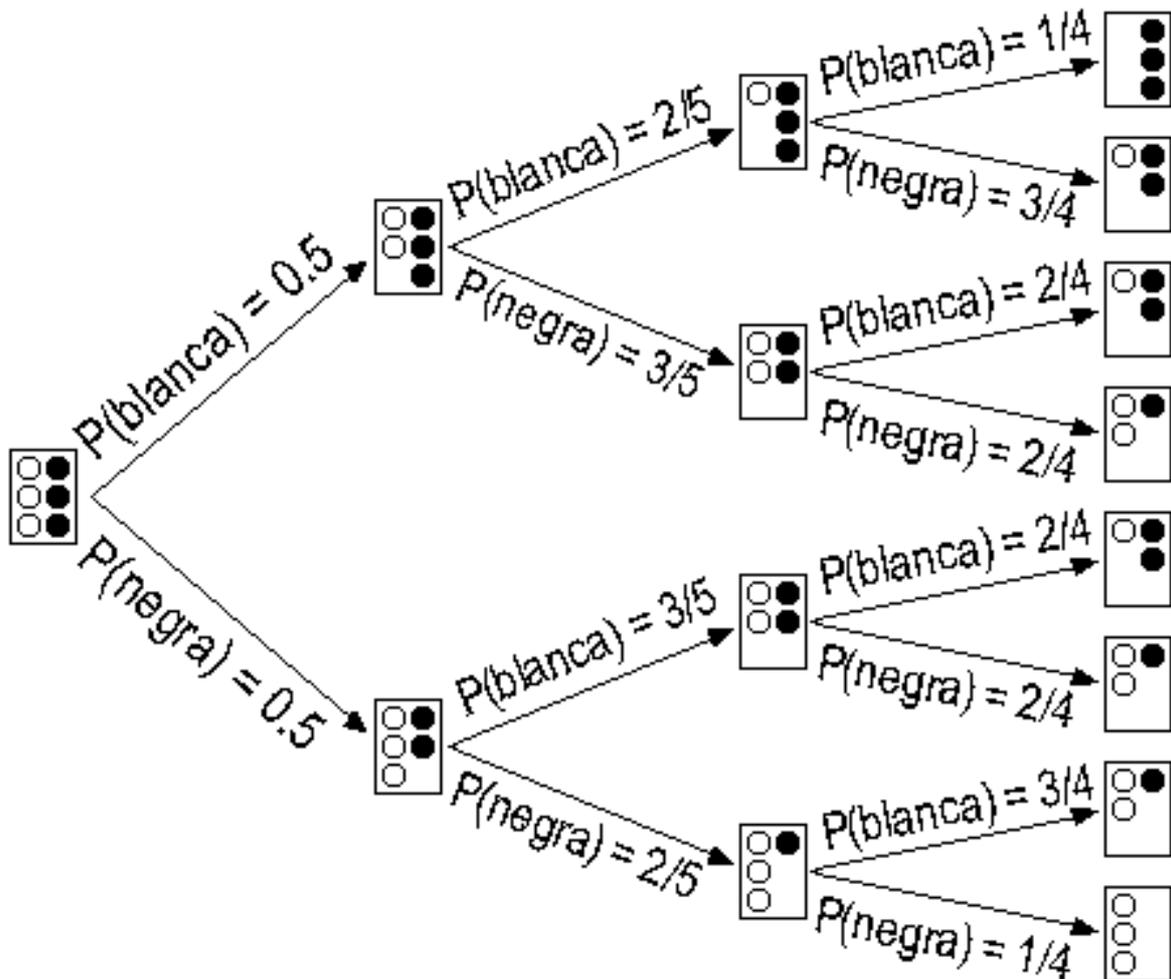
muestra el estado original de la caja, las

probabilidades de sacar una bolita blanca y una bolita negra, y el estado de la caja luego de sacar ese tipo de bolita.

Naturalmente, el diagrama se puede expandir, y se

puede volver a describir las probabilidades de sacar

blancas y negras en cada caso (es decir, las la segunda bolita que se saque sea blanca o negra) y



Este gráfico es una versión ampliada del anterior. Para cada situación hipotética, se volvió a calcular la probabilidad de sacar una bolita blanca o negra, y se volvió a dibujar el estado en que quedaría la caja si sucediera que se extrajera una bolita de ese color. A medida que vamos recorriendo los caminos va cambiando el dibujo de la cajita; esto lo que muestra es que va cambiando el espacio muestral a medida que vamos sacando bolitas. Es por eso que las probabilidades que aparecen en las flechas son condicionales, referidas al espacio muestral del que parte cada flecha.

Este diagrama nos proporciona muchísima información. Por ejemplo:

Podemos calcular fácilmente lo que habíamos calculado antes: la probabilidad de que las primeras 2 que se saquen sean blancas y la tercera negra. Simplemente hacemos el camino correspondiente, multiplicando, y obtenemos la probabilidad buscada:

$$0.5 \cdot \frac{2}{5} \cdot \frac{3}{4} = 0.15$$

Pero este es sólo uno de los 8 caminos posibles. Todos se pueden calcular de la misma forma.

No es solamente la probabilidad de los caminos de 3 bolitas la que podemos calcular. También podemos usar el diagrama para calcular las probabilidades de los caminos de 2 bolitas. Por ejemplo, la probabilidad de sacar primero 1 blanca y después 1 negra es: $0.5 \cdot \frac{3}{5} = \frac{3}{10}$

Todos esos cálculos los podemos hacer porque las probabilidades que figuran en el diagrama son, en realidad, probabilidades condicionales. Por ejemplo, arriba a la derecha dice "P(negra) = $\frac{3}{4}$ ". Si los sucesos A, B y C son como los definimos antes, esa probabilidad que aparece en el gráfico no es sino $P(C / (A \cap B))$. Es decir, el "P(negra) = $\frac{3}{4}$ " que aparece en el gráfico significa "la probabilidad de que la tercera bolita extraída sea negra, dado que las dos primeras fueron blancas, es $\frac{3}{4}$ ".

Otro tipo de cálculo que nos podría interesar hacer es: "¿cuál es la probabilidad de que luego de sacar 3 bolitas, queden dentro de la caja 2 negras y 1 blanca?". Para calcular esta probabilidad, primero hay que buscar todos los caminos que nos conducen a esa situación:

$$C_1 = B, B, N \quad C_2 = B, N, B \quad C_3 = N, B, B$$

Luego, la probabilidad de terminar teniendo en la caja 2 negras y 1 blanca es la probabilidad de haber hecho el camino 1 ó el camino 2 ó el camino 3, es decir:

$$P(C_1 \cup C_2 \cup C_3)$$

Como los caminos son disjuntos (porque si se hace uno, es imposible que se hagan los otros), entonces la probabilidad de la unión es la suma de las probabilidades, con lo cual:

$$P(C_1 \cup C_2 \cup C_3) = P(C_1) + P(C_2) + P(C_3)$$

Y usando el diagrama para calcular las probabilidades, obtenemos:

$$P(C_1) + P(C_2) + P(C_3) = 0.5 \cdot \frac{2}{5} \cdot \frac{3}{4} + 0.5 \cdot \frac{3}{5} \cdot \frac{2}{4} + 0.5 \cdot \frac{3}{5} \cdot \frac{2}{4} = \frac{9}{20}$$

Además notemos que:

- en todas las bifurcaciones, $P(\text{blanca}) + P(\text{negra}) = 1$, porque si sacamos una bolita, tendrá necesariamente que ser blanca o negra. No hay ninguna otra posibilidad.
- si sumamos las probabilidades de efectuar cada uno de los 8 caminos que tenemos si sacamos 3 bolitas, esa suma debe dar 1, porque si sacamos 3 bolitas, tendremos necesariamente que emplear uno de los 8 caminos. No hay ninguna otra posibilidad. Esto también se cumple para los caminos que resultan de sacar 2 bolitas, y para los que resultan de sacar 1 bolita.

Por último, recordemos los gráficos sirven para mostrar, no para justificar. Si se nos pide una justificación, se requiere el tipo de análisis que hemos hecho "formalmente".

Aplicando dos veces la definición de probabilidad condicional

La definición de probabilidad condicional es:

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

Pero como $P(A \cap B) = P(B \cap A)$ y además:

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

Combinando las dos expresiones resulta:

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

Es decir:

$$P(A/B) = \frac{P(B/A) P(A)}{P(B)} \quad \text{o bien} \quad P(B/A) = \frac{P(A/B) P(B)}{P(A)}$$

lo cual puede resultarnos útil si tenemos $P(A/B)$ y queremos conocer $P(B/A)$ o viceversa. Recordemos que no son lo mismo.

Ejemplo

El 30% de las personas tiene ojos claros. El 60% de las personas es mujer. Se sabe además que la probabilidad de que una mujer tenga ojos claros es 0,2. ¿Cuál es la probabilidad de que una persona de ojos claros sea mujer?

Trabajaremos con los sucesos:

A: la persona extraída tiene ojos claros

B: la persona extraída es mujer

Entonces los datos son:

$$P(A) = 0,3$$

$$P(B) = 0,6$$

$$P(A/B) = 0,2$$

Y queremos saber $P(B/A)$. Usando el resultado anterior obtenemos:

$$P(B/A) = \frac{P(A/B) P(B)}{P(A)} = \frac{0,2 \cdot 0,6}{0,3} = 0,4$$

Problemas típicos

1) Se tiene que: $P(A) = 0.3$, $P(A/B) = 0.4$, $P(A \cap B) = 0.2$. Calcule $P(B)$ y $P(B/A)$.

Resolución:

Por la definición de probabilidad condicional, $P(A/B) = \frac{P(A \cap B)}{P(B)}$.

Despejando $P(B)$, queda: $P(B) = \frac{P(A \cap B)}{P(A/B)}$. Luego $P(B) = \frac{0.2}{0.4} = 0.5$.

Nuevamente, por la definición de probabilidad condicional,

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{0.2}{0.3} = 0.67$$

2) La probabilidad de que llueva en un determinado día es 0.4. Pero si la tribu baila la danza de la lluvia, la probabilidad de que llueva se duplica. En la aldea tienen la costumbre de bailar la danza de la lluvia todos los días, a menos que hayan salido a cazar rinocerontes. La tribu sale a cazar rinocerontes el 70% de los días. Calcule la probabilidad de que en un determinado día:

- a) llueva
- b) llueva, sabiendo que la tribu bailó la danza de la lluvia
- c) la tribu baile la danza de la lluvia
- d) llueva y la tribu baile la danza de la lluvia
- e) la tribu haya bailado la danza de la lluvia, dado que ese día terminó lloviendo
- f) la tribu baile la danza de la lluvia y no llueva
- g) llueva, sabiendo que ese día la tribu no baila la danza de la lluvia

Resolución:

Comencemos por definir, para un día cualquiera:

A: llueve

B: la tribu baila la danza de la lluvia

Los datos que nos dan son:

$$P(A) = 0.4$$

$$P(A/B) = 0.8$$

$$P(B) = 0.3 \text{ (porque el 70\% de los días la tribu está fuera de la aldea cazando rinocerontes)}$$

- a) La probabilidad de que llueva es dato, $P(A) = 0.4$
- b) La probabilidad de que llueva, sabiendo que la tribu bailó la danza de la lluvia, también es dato. $P(A/B) = 0.8$
- c) La probabilidad de que la tribu baile la danza de la lluvia es, como calculamos antes, $P(B) = 0.3$
- d) La probabilidad de que llueva y la tribu baile la danza de la lluvia es, por la definición de probabilidad condicional, $P(A \cap B) = P(A/B) \cdot P(B) = 0.24$

e) La probabilidad de que la tribu haya bailado la danza de la lluvia, dado que ese día terminó lloviendo, es $P(B/A)$. Obtenemos:

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{0.24}{0.4} = 0.6$$

f) La probabilidad de que en un determinado día la tribu baile la danza de la lluvia y no llueva, es $P(B \cap A^c)$

Por propiedades de conjuntos, sabemos que $P(B \cap A) + P(B \cap A^c) = P(B)$, porque $(B \cap A) \cup (B \cap A^c) = B$. Esto también puede entenderse como que la probabilidad de que la tribu baile y llueva, más la probabilidad de que la tribu baile y no llueva, es la probabilidad de que la tribu baile (sin importar si termina lloviendo o no). Mediante cualquiera de las dos justificaciones, $P(B \cap A^c) = P(B) - P(B \cap A)$, con lo cual la probabilidad pedida es $P(B) - P(B \cap A) = 0.06$

Vemos que este resultado es coherente, ya que de acuerdo a los datos, la danza de la lluvia suele ser bastante efectiva.

g) La probabilidad de que llueva, sabiendo que ese día la tribu había salido a cazar rinocerontes, y por lo tanto no bailó la danza de la lluvia, es $P(A/B^c)$, es decir, "probabilidad de A dado que no B". Por el teorema de la probabilidad condicional, queda:

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

Por propiedades de conjuntos, sabemos que $P(A \cap B) + P(A \cap B^c) = P(A)$, porque $(A \cap B) \cup (A \cap B^c) = A$. Esto también puede entenderse como que la probabilidad de que llueva y la tribu baile, más la probabilidad de que llueva y la tribu no baile, es la probabilidad de que llueva (sin importar si la tribu baila o no).

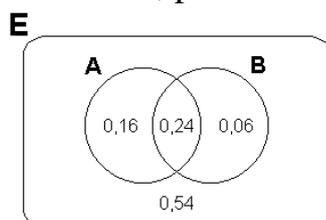
Entonces $P(A \cap B^c) = P(A) - P(A \cap B)$, con lo cual: $P(A/\bar{B}) = \frac{P(A) - P(A \cap B)}{P(\bar{B})}$

Además sabemos que $P(B) + P(B^c) = 1$, con lo cual queda:

$$P(A/\bar{B}) = \frac{P(A) - P(A \cap B)}{1 - P(B)}$$

Y ya dejamos todo en función de valores que ya conocemos. Hacemos la cuenta y obtenemos que $P(A/B^c) = 0.23$

Por último, podríamos hacer un gráfico para visualizar todo más claramente:



Primero colocamos en la intersección que $P(A \cap B) = 0.24$

Luego, como $P(A) = 0.4$, entonces $P(A \cap B^c)$ debe ser 0.16, para satisfacer $P(A \cap B) + P(A \cap B^c) = P(A)$.

Análogamente, como $P(B) = 0.3$, entonces $P(B \cap A^c)$ debe ser 0.06, para satisfacer $P(B \cap A) + P(B \cap A^c) = P(B)$.

Por otro lado, sabemos que la probabilidad total, es decir, la probabilidad de E, debe ser 1. Como la probabilidad total es 1, deducimos el valor que nos falta, es decir, la probabilidad de que no suceda ni A ni B. $P(A^c \wedge B^c)$ vale $1 - 0.16 - 0.24 - 0.06 = 0.54$

Si solamente hubiéramos querido las respuestas a las preguntas de este problema, podríamos haber hecho el gráfico, completado con los datos, y obtener las respuestas rápidamente. Hicimos el análisis expuesto para mostrar una posible justificación de los resultados obtenidos. Recordemos que no hay una única forma de aplicar la probabilidad condicional para llegar al resultado, y también que los gráficos no constituyen una justificación.

3) En una determinada ciudad, el 11% de las personas tiene el cabello rubio y el 89% tiene el cabello negro. En esa población, 49 de cada 100 personas son hombres. Tomando una persona al azar, existe una probabilidad 0.84 de que esa persona tenga ojos oscuros. El 54.55% de las personas rubias, también tienen ojos claros. El 13.73% de las mujeres son rubias. El 42% de las personas son hombres de ojos oscuros. El 41% de las personas no es mujer ni tiene cabello rubio ni ojos claros.

Calcule la probabilidad de una persona tomada al azar:

- a) Sea una mujer rubia de ojos claros.**
- b) Tenga cabello negro y ojos claros.**
- c) Sea un hombre rubio de ojos oscuros.**

Resolución:

Comencemos definir los sucesos y organizar los datos:

R: que una persona sea rubia

C: que una persona tenga ojos claros

M: que una persona sea mujer

$$P(R) = 0.11$$

$$P(M) = 0.51$$

$$P(C) = 0.16$$

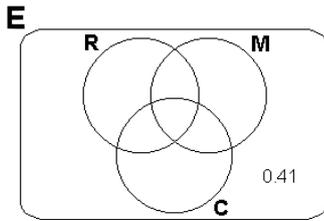
$$P(C/R) = 0.5455$$

$$P(R/M) = 0.1317$$

$$P(C^c \wedge M^c) = 0.42$$

$$P(C^c \wedge R^c \wedge M^c) = 0.41$$

Como el problema es complicado, conviene que hagamos un gráfico y vayamos completando los valores a medida que los obtenemos:

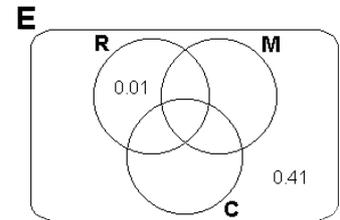


Vemos que, con 3 sucesos, E queda dividido en $2^3 = 8$ regiones. De las 8 regiones, el único dato que conocemos que abarca a una sola región es $P(C^c \wedge R^c \wedge M^c) = 0.41$

Por propiedades de conjuntos, como vimos en los ejemplos anteriores,

$$P(C^c \wedge M^c) = P(C^c \wedge M^c \wedge R) + P(C^c \wedge M^c \wedge R^c)$$

con lo cual $P(C^c \wedge M^c \wedge R) = 0.01$. El gráfico queda:



Ahora observemos que en R hay 4 regiones, y también tenemos 4 datos:

$$P(R) = 0.11$$

$$P(C^c \wedge M^c \wedge R) = 0.01$$

$$P(C/R) = 0.5455$$

$$P(R/M) = 0.1317$$

De las dos condicionales podemos obtener:

$$P(C \wedge R) = P(C/R) \cdot P(R) = 0.06$$

$$P(R \wedge M) = P(R/M) \cdot P(M) = 0.07$$

Podemos escribir a R como $R = R \cap C \cap M \cup R \cap C^c \cap M \cup R \cap C \cap M^c \cup R \cap C^c \cap M^c$

Por propiedades de conjuntos, $R \cap C \cap M \cup R \cap C \cap M^c = R \cap C$, con lo cual

$$R = R \cap C \cup R \cap C^c \cap M \cup R \cap C^c \cap M^c$$

Con lo cual $P(R) = P(R \cap C \cup R \cap C^c \cap M \cup R \cap C^c \cap M^c)$

Como esos 3 subconjuntos de R son disjuntos, entonces:

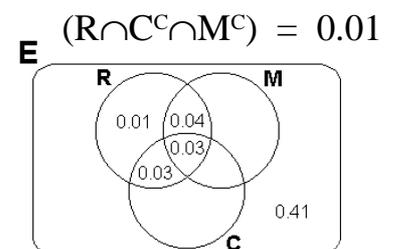
$$P(R) = P(R \cap C) + P(R \cap C^c \cap M) + P(R \cap C^c \cap M^c)$$

Y sabemos que $P(R) = 0.11$, $P(R \cap C) = 0.06$ y P

Por lo tanto, $P(R \cap C^c \cap M) = 0.04$

Y luego $P(R \cap C \cap M) = P(R \cap M) - P(R \cap C^c \cap M) = 0,03$

Con lo cual encontramos la probabilidad que nos pedían en a), y el gráfico nos queda:



Ahora vamos a aplicar la fórmula para la suma de 3 sucesos:

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

En este caso queda:

$$P(R) + P(M) + P(C) - P(R \cap M) - P(R \cap C) - P(M \cap C) + P(R \cap M \cap C)$$

Y sabemos que esta suma da $1 - 0.41 = 0.59$

El único valor que no conocemos es $P(M \cap C)$. Lo despejamos:

$$P(M \cap C) = P(R) + P(M) + P(C) - P(R \cap M) - P(R \cap C) + P(R \cap M \cap C) - 0.59$$

$$P(M \cap C) = 0.11 + 0.51 + 0.16 - 0.06 - 0.07 + 0.03 - 0.59$$

$$P(M \cap C) = 0.09$$

$$\text{De donde obtenemos } P(M \cap C \cap R^c) = P(M \cap C) - P(M \cap C \cap R) = 0.06$$

Y ya podemos obtener directamente los dos valores que faltan para completar el gráfico:

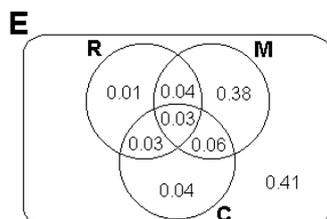
$$P(C \cap M^c \cap R^c) = P(C) - P(C \cap M^c \cap R) - P(C \cap M \cap R) - P(C \cap M \cap R^c)$$

$$P(C \cap M^c \cap R^c) = 0.16 - 0.03 - 0.03 - 0.06 = 0.04$$

$$P(M \cap C^c \cap R^c) = P(M) - P(M \cap C^c \cap R) - P(M \cap C \cap R) - P(M \cap C \cap R^c)$$

$$P(M \cap C^c \cap R^c) = 0.38 - 0.04 - 0.03 - 0.06 = 0.38$$

El gráfico queda:



En b) nos piden la probabilidad de que una persona tenga cabello negro y ojos claros. Eso es $P(C \cap R^c) = P(C \cap R^c \cap M) + P(C \cap R^c \cap M^c) = 0.06 + 0.04 = 0.1$

En c) nos piden $P(R \cap C^c \cap M^c) = 0.01$

4) En una ciudad hay un 60% de mujeres. El 25% de las personas tiene ojos claros. El 30% de las mujeres tiene ojos claros. ¿Qué porcentaje de los hombres tiene ojos oscuros?

Resolución:

Podríamos tomar los eventos:

M: ser mujer

C: tener ojos claros

Y proceder exactamente igual que en el problema 2, en cuyo caso estaríamos buscando: $P(C^c / M^c)$

Sin embargo, vamos a hacer un planteo un poco diferente, con el objeto de mostrar un método más general para un tipo de problema que veremos más adelante.

Tomaremos los eventos:

M: ser mujer

H: ser hombre

C: tener ojos claros

O: tener ojos oscuros

Y estamos buscando: $P(O / H)$

Comenzaremos por hacer el siguiente esquema:

	M	H	
C			
O			

En las 4 posiciones centrales, colocaremos las probabilidades de los 4 casos posibles ($M \cap C$, $M \cap O$, $H \cap C$, $H \cap O$). En las 4 posiciones periféricas, colocaremos las probabilidades propias de cada uno de los sucesos M, H, C y O. Más adelante estudiaremos estas últimas probabilidades con el nombre de "probabilidades marginales".

Veamos qué datos nos dan:

$$P(M) = 0,6$$

$$P(C) = 0,25$$

$$P(C/M) = 0,3$$

Nos piden:

$$P(O/H)$$

Agregamos los 2 primeros datos al gráfico:

	M	H	
C			0.25
O			
	0.6		

Observemos que los hemos colocado en las posiciones periféricas.

Además, como sabemos que una persona sólo puede tener ojos claros u ojos oscuros, entonces $P(C) + P(O) = 1$, con lo cual $P(O) = 0.75$. Análogamente, $P(M) + P(H) = 1$, con lo cual $P(H) = 0.4$.

	M	H	
C			0.25
O			0.75
	0.6	0.4	

El tercer dato nos dice:

$$P(C/M) = 0,3$$

Por la definición de probabilidad condicional, podemos escribir: $P(C/M) = \frac{P(C \cap M)}{P(M)} = 0.3$

, con lo cual: $P(C \cap M) = P(M) \cdot P(C/M) = 0,6 \cdot 0,3 = 0,18$

Eso nos da el valor de una de las intersecciones, y las intersecciones son las que están en el centro del gráfico. Coloquemos el valor que acabamos de obtener:

	M	H	
C	0.18		0.25

O			0.75
	0.6	0.4	

Notemos que el porcentaje de personas con ojos claros será el porcentaje de mujeres con ojos claros más el porcentaje de hombres con ojos claros. Entonces:

$$P(C) = P(C \cap M) + P(C \cap H)$$

Con lo cual:

$$P(C \cap H) = P(C) - P(C \cap M) = 0.25 - 0.18 = 0.07$$

Se verá un mejor acercamiento a ese planteo, en la sección "probabilidad total" de este mismo capítulo.

Análogamente, el porcentaje de mujeres será el porcentaje de mujeres con ojos claros más el porcentaje de mujeres con ojos oscuros. Entonces:

$$P(M) = P(M \cap C) + P(M \cap O)$$

Con lo cual:

$$P(M \cap O) = P(M) - P(M \cap C) = 0.6 - 0.18 = 0.42$$

Completando el gráfico:

	M	H	
C	0.18	0.07	0.25
O	0.42	0.33	0.75
	0.6	0.4	

Sólo resta aplicar una vez más el mismo razonamiento análogo para calcular todas las probabilidades. Podemos plantear "el porcentaje de hombres es el porcentaje de hombres con ojos claros, más el porcentaje de hombres con ojos oscuros" o bien "el porcentaje de personas con ojos oscuros es el porcentaje de mujeres con ojos oscuros más el porcentaje de hombres con ojos oscuros".

Si hacemos el primero, nos queda:

$$P(H) = P(H \cap C) + P(H \cap O)$$

Con lo cual:

$$P(H \cap O) = P(H) - P(H \cap C) = 0.4 - 0.07 = 0.33$$

También sabemos que como los 4 casos $M \cap C$, $M \cap O$, $H \cap C$, $H \cap O$ son todos los casos posibles, entonces la suma de sus probabilidades debe dar 1.

$$P(M \cap C) + P(M \cap O) + P(H \cap C) + P(H \cap O) = 1$$

Con lo cual:

$$P(H \cap O) = 1 - P(M \cap C) - P(M \cap O) - P(H \cap C) = 1 - 0.18 - 0.42 - 0.07 = 0.33$$

$$\text{Luego } P(O|H) = \frac{P(O \cap H)}{P(H)} = \frac{0.33}{0.4} = 0.825$$

Lo cual indica que el 82,5% de los hombres tiene ojos oscuros.

5) Las revistas pueden estar en castellano, en inglés o en portugués. En cierto puesto de diarios, el 90% de las revistas está en castellano y el 2% está en portugués. El 80% de las revistas de computación está en castellano. El 30% de las revistas es de computación. Si una revista está en portugués, hay una probabilidad 0,4 de que sea de computación. ¿Cuál es la probabilidad de que tomando una revista al azar, esté en inglés y no sea de computación?

Resolución:

Observemos que este problema es como el anterior, pero en vez de ser de 2x2 es de 3x2. Planteamos el mismo tipo de diagrama que en el ejemplo anterior:

	Castellano	Inglés	Portugués	
Computación				
No comput.				

Los datos son:

$$P(\text{castellano}) = 0.9$$

$$P(\text{portugués}) = 0.02$$

$$P(\text{computación}) = 0.3$$

$$P(\text{castellano} / \text{computación}) = 0.8$$

$$P(\text{computación} / \text{portugués}) = 0.4$$

Y nos piden:

$$P(\text{inglés} \wedge \text{computación})$$

Colocamos los 3 primeros datos:

	Castellano	Inglés	Portugués	
Computación				0.3
No comput.				
	0.9		0.02	

También sabemos que $P(\text{no comput}) = 0.7$ porque como es el negado de un suceso, su probabilidad es 1 menos la probabilidad del suceso. Visto de otra forma, $P(\text{computación}) + P(\text{no computación}) = 1$.

Análogamente, como las revistas tienen que estar necesariamente en uno de los 3 idiomas, entonces $P(\text{castellano}) + P(\text{inglés}) + P(\text{portugués}) = 1$

$$\Rightarrow P(\text{inglés}) = 1 - 0.9 - 0.02 = 0.08$$

	Castellano	Inglés	Portugués	
Computación				0.3
No comput.				0.7
	0.9	0.08	0.02	

Por el teorema de la probabilidad condicional:

$$P(\text{castellano} / \text{computación}) = 0.8 \Rightarrow P(\text{castellano} \cap \text{computación}) = 0.8 \cdot 0.3 = 0.24$$

$$P(\text{computación} / \text{portugués}) = 0.4 \Rightarrow P(\text{computación} \cap \text{portugués}) = 0.4 \cdot 0.02 = 0.008$$

	Castellano	Inglés	Portugués	
Computación	0.24		0.008	0.3
No comput.				0.7
	0.9	0.08	0.02	

Luego, por las propiedades ya estudiadas:

$$P(\text{castellano}) = P(\text{castellano} \cap \text{computación}) + P(\text{castellano} \cap \text{no computación})$$

$$\Rightarrow P(\text{castellano} \cap \text{no computación}) = 0.9 - 0.24 = 0.66$$

Y análogamente completamos el resto del cuadro.

	Castellano	Inglés	Portugués	
Computación	0.24	0.052	0.008	0.3
No comput.	0.66	0.028	0.012	0.7
	0.9	0.08	0.02	

Luego la probabilidad pedida es $P(\text{inglés} \cap \text{computación}) = 0.052$

6) En una caja hay 40 bolitas: 10 negras, 10 blancas, 10 rojas y 10 verdes. Se sacan 4 bolitas (sin reposición). ¿Cuál es la probabilidad de que las 4 bolitas extraídas sean de colores distintos?

Resolución:

Hay 2 formas de resolver este problema: una es mediante multiplicando probabilidades, como ya se vio, y la otra mediante la definición de Laplace y el cálculo combinatorio, como se verá más adelante. Aquí resolveremos el problema de la primera forma.

Podemos pensar el problema así: sacar 4 bolitas de colores distintos, es como sacar primero una bolita cualquiera (no importa el color), y luego sacar una segunda bolita (que sea de color distinto a la primera), y luego que la tercera sea de color distinto a la primera y la segunda, y luego que la cuarta sea de color distinto a las 3 primeras. Podríamos tomar los siguientes sucesos:

A: que cuando saque la segunda, el color sea distinto al de la primera.

B: que cuando saque la tercera, el color sea distinto al de la primera y la segunda.

C: que cuando saque la cuarta, el color sea distinto al de las 3 primeras.

Y luego podemos buscar $P(A \cap B \cap C)$. Si llamamos $D = A \cap B$, entonces podemos escribir:

$$P(A \cap B \cap C) = P(C \wedge D) = P(C / D) \cdot P(D)$$

$$P(D) = P(B \cap A) = P(B / A) \cdot P(A)$$

Ahora analicemos:

Si hay 10 bolitas de cada color, entonces sin importar de qué color sea la primera que saquemos, quedarán 9 del mismo color, y 30 de otros colores. Entonces cuando saquemos la segunda bolita, nos quedarán 30 bolitas favorables, entre un total de 39 bolitas.

Entonces la probabilidad de que la segunda bolita sea de un color distinto al de la primera es $P(A) = \frac{30}{39}$.

Luego sacar la segunda bolita, y suponiendo que fue de un color distinto al de la primera, nos quedarán en la caja 38 bolitas, de las cuales 9 serán del color de la primera, 9 serán del color de la segunda y 20 serán de los 2 colores que todavía no salieron. Entonces la probabilidad de que la tercera bolita sea de color distinto al de las 2 primeras, suponiendo que las 2 primeras fueron de colores distintos, no es otra cosa que la probabilidad de B dado A, y como quedan 20 bolitas favorables de en un total de 38, vale $P(B / A) = \frac{20}{38}$.

Con esto ya hemos calculado $P(D)$, porque según habíamos determinado antes, valía:

$$P(D) = P(B / A) \cdot P(A) = \frac{20}{38} \cdot \frac{30}{39}$$

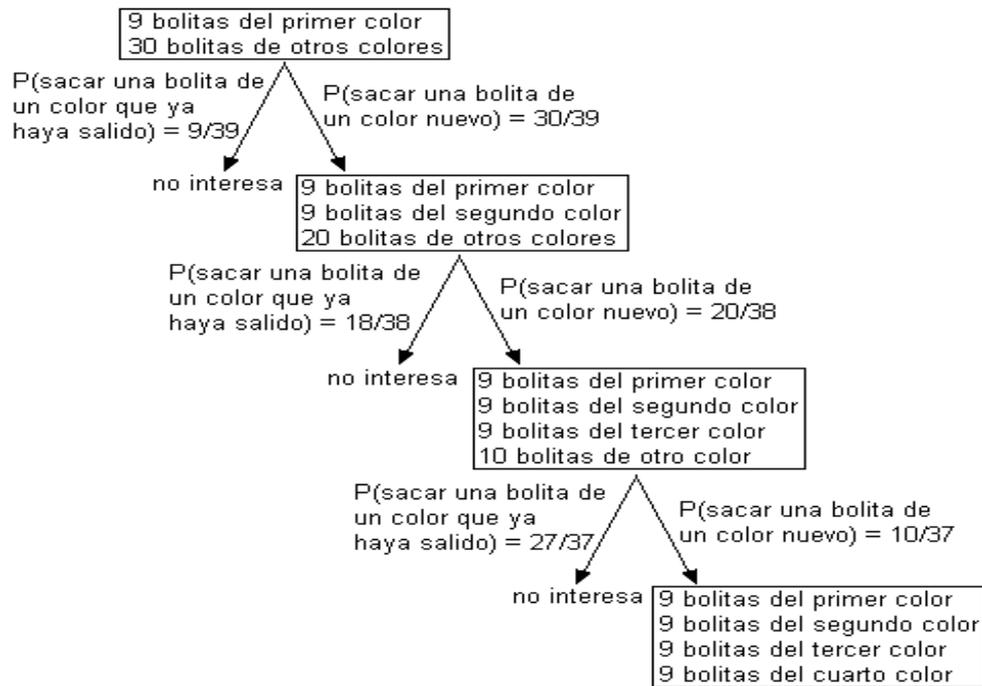
Usando el mismo razonamiento anterior, si suponemos que las primera 3 bolitas extraídas fueron de distintos colores, entonces quedan 37 bolitas, de las cuales 9 son del color de la primera, 9 del color de la segunda, 9 del color de la tercera, y 10 del color que no salió.

Entonces la probabilidad de que la cuarta bolita sea de color distinto al de las 3 primeras, suponiendo que las 3 primeras fueron de colores distintos, no es otra cosa que la probabilidad de C dado D, y como quedan 10 bolitas favorables de en un total de 37, vale $P(C / D) = \frac{10}{37}$.

Y con esto ya hemos calculado $P(A \cap B \cap C)$, porque según habíamos determinado antes, valía:

$$P(A \cap B \cap C) = P(C / D) \cdot P(D) = \frac{30}{39} \cdot \frac{20}{38} \cdot \frac{10}{37} = 0.10942$$

También, sin salirnos del modelo de sucesos sucesivos, podemos pensar el problema como un árbol. Sacamos la primera bolita (de cualquier color) y queda:



Y así, multiplicamos $\frac{30}{39} \cdot \frac{20}{38} \cdot \frac{10}{37}$, con lo cual obtenemos el mismo resultado. Por último, recordemos que un gráfico sólo sirve para mostrar información, no para justificarla. Para justificar este resultado, debemos emplear probabilidad condicional.

7) Se tienen en una urna 2 bolas negras, 3 blancas y 4 rojas. Calcule la probabilidad de que al sacar 3 bolas sin reposición

- sean 3 blancas
- la primera sea blanca, la segunda negra, y la tercera roja
- sea una de cada color

Resolución:

a)

A: la primera bola es blanca

B: la segunda bola es blanca

C: la tercera bola es blanca

Se pide: $P(A \cap B \cap C)$

Lo cual como vimos antes se puede escribir como:

$$P(A \cap B \cap C) = P(A) P\left(\frac{B}{A}\right) P\left(\frac{C}{A \cap B}\right)$$

Análogamente a como procedimos antes:

Tenemos 9 bolas (2 negras, 3 blancas, 4 rojas)

Luego $P(A) = 3/9$

Si sacamos una blanca (es decir, nos metemos en el espacio muestral en el cual se asume que se sacó una bola blanca) tenemos 8 bolas (2 negras, 2 blancas, 4 rojas)

Luego $P(B/A) = 2/8$

Si sacamos otra blanca (es decir, nos metemos en el espacio muestral en el cual se asume que se sacaron dos bolas blancas) tenemos 7 bolas (2 negras, 1 blanca, 4 rojas)

Luego $P(C / A \cap B) = 1/7$

Luego, $P(A \cap B \cap C) = 6/504 = 0,0119$

b)

Este ejercicio es muy similar al anterior. Planteamos:

A: la primera bola es blanca

B: la segunda bola es negra

C: la tercera bola es roja

Se pide: $P(A \cap B \cap C)$

$$P(A \cap B \cap C) = P(A) P\left(\frac{B}{A}\right) P\left(\frac{C}{A \cap B}\right)$$

Análogamente a como procedimos antes:

Tenemos 9 bolas (2 negras, 3 blancas, 4 rojas)

Luego $P(A) = 3/9$

Si sacamos una blanca (es decir, nos metemos en el espacio muestral en el cual se asume que se sacó una bola blanca) tenemos 8 bolas (2 negras, 2 blancas, 4 rojas)

Luego $P(B/A) = 2/8$

Si sacamos una negra (es decir, nos metemos en el espacio muestral en el cual se asume que se sacaron una blanca y una negra) tenemos 7 bolas (1 negra, 2 blancas, 4 rojas)

Luego $P(C / A \cap B) = 4/7$

Luego, $P(A \cap B \cap C) = 24/504 = 1/21$

c)

Si pensamos este problema como un árbol de los que vimos antes, tenemos un diagrama en el cual de cada punto salen 3 opciones (negra, blanca, roja). Si vamos a considerar las formas posibles de sacar 3 bolitas, tendremos $3 \cdot 3 \cdot 3 = 3^3 = 27$ formas posibles.

Las formas posibles de sacar 3 bolitas de distintos colores son $3 \cdot 2 \cdot 1 = 6$ (primero tenemos 3 colores disponibles, luego 2, luego sólo 1). Entonces la probabilidad que nos piden es la suma de 6 caminos, de los 27 que el árbol tiene en total. Vemos que lo que nos pedían en a) y en b) eran simplemente 2 caminos de los 27 que hay.

Por lo tanto una de las formas de hallar la probabilidad pedida en c) (más adelante veremos otras) es sumando 6 ramas del árbol, cada una de las cuales se obtiene como en los dos puntos anteriores. Entonces:

$$P(\text{negra, luego blanca, luego roja}) = 2/9 \cdot 3/8 \cdot 4/7 = 24/504 = 1/21$$

$$P(\text{negra, luego roja, luego blanca}) = 2/9 \cdot 4/8 \cdot 3/7 = 1/21$$

$$P(\text{blanca, luego negra, luego roja}) = 3/9 \cdot 2/8 \cdot 4/7 = 1/21$$

$$P(\text{blanca, luego roja, luego negra}) = 3/9 \cdot 4/8 \cdot 2/7 = 1/21$$

$$P(\text{roja, luego blanca, luego negra}) = 4/9 \cdot 3/8 \cdot 2/7 = 1/21$$

$$P(\text{roja, luego negra, luego blanca}) = 4/9 \cdot 2/8 \cdot 3/7 = 1/21$$

Luego la respuesta es $6/21 = 2/7$

No deja de ser llamativo que las 6 ramas hayan dado lo mismo. Esto es porque en realidad el problema puede ser visto de forma mucho más simple. Dicha forma será estudiada más adelante. Pero esta solución se ofrece porque es mecánica, funciona siempre, se puede programar, y no da lugar a equivocaciones.

Independencia

Dos sucesos son independientes si el hecho de conocer que ocurrió uno de ellos no afecta la probabilidad de que ocurra el otro.

Consideremos por ejemplo los siguientes sucesos:

A: Argentina le gana hoy a Brasil en el partido de fútbol

B: Esta noche hay luna llena

C: Sube el precio de los autos nuevos

D: Se reduce la cantidad de gente que compra autos nuevos

Dijimos que dos sucesos son independientes si el hecho de conocer que ocurrió uno de ellos no afecta la probabilidad de que ocurra el otro.

Hoy Argentina y Brasil jugarán un partido de fútbol, y con nuestro conocimiento futbolístico llegamos a la conclusión de que la probabilidad de que Argentina le gane hoy a Brasil es de 0,6.

En ese momento miramos por la ventana y nos damos cuenta de que hoy hay luna llena. ¿Eso modificará nuestra creencia de que la probabilidad de que Argentina le gane a Brasil es 0,6? Es decir, la probabilidad de que gane Argentina en una noche que hay luna llena, ¿podríamos decir que es distinta de la probabilidad de que gane Argentina en una noche cualquiera? Probablemente no, a menos que seamos expertos en astrología y “sepamos” que los astros afectan el desempeño de los futbolistas de distintos países.

Dicho de otra forma, $P(A) = 0,6$ y además $P(A/B) = 0,6$ (porque el hecho de saber que ocurrió B no afecta la probabilidad de que ocurra A).

Vemos que $P(A) = P(A/B)$ es una forma matemática de expresar lo que dijimos antes de que dos sucesos son independientes si el hecho de conocer que ocurrió uno de ellos no afecta la probabilidad de que ocurra el otro.

Supongamos que la historia hubiera sido distinta: Sabemos que la cuarta parte de los días hay luna llena, y entonces $P(B) = 0,25$. Si alguien nos pregunta: “¿cuál es la probabilidad de que el 26 de abril de 1982 haya habido luna llena?”, responderemos: “0,25”. Luego la persona nos dice: “¿Estás seguro? Mirá que ese día Argentina le ganó a Brasil”.

¿Modificaremos entonces nuestra respuesta? Probablemente no, a menos que a la luna le guste ponerse llena cuando Argentina le gana a Brasil.

Dicho de otra forma, $P(B) = 0,25$ y además $P(B/A) = 0,25$ (porque el hecho de saber que Argentina le ganó a Brasil no afecta la probabilidad de que haya habido luna llena). *Observamos entonces que en este ejemplo también vale $P(B) = P(B/A)$. Y si hacemos las correspondientes cuentas, también veremos que se verifica $P(A \cap B) = P(A) \cdot P(B)$* Daremos a continuación la definición y luego demostraremos las equivalencias:

<p>Dos sucesos A, B son independientes</p> <p>\Leftrightarrow</p> <p>$P(A/B) = P(A)$</p> <p>\Leftrightarrow</p> <p>$P(B/A) = P(B)$</p> <p>\Leftrightarrow</p> <p>$P(A \cap B) = P(A) \cdot P(B)$</p>
--

Verificaremos las equivalencias:

Si se cumple **$P(A/B) = P(A)$** , aplicamos la definición de probabilidad condicional del lado izquierdo y nos queda: $P(A \cap B) / P(B) = P(A)$, luego **$P(A \cap B) = P(A) \cdot P(B)$**

Si pensamos el $P(A \cap B)$ como $P(B \cap A)$ y aplicamos nuevamente la definición de probabilidad condicional del lado izquierdo, nos queda $P(B/A) \cdot P(A) = P(A) \cdot P(B)$, luego **$P(B/A) = P(B)$** , con lo cual verificamos la equivalencia de las 3 expresiones.

Pasando a los sucesos C y D, aún sin saber mucho de economía nos imaginamos que debe haber una cierta relación entre los precios y la cantidad de compradores. No nos resultaría extraño que la probabilidad de que se reduzca la cantidad de compradores de autos nuevos en un país donde ha aumentado el costo de los autos nuevos sea mayor que en un país cualquiera en el cual no sabemos si aumentó o no aumentó el costo de los autos nuevos. Supongamos que del anuario de la sociedad internacional de automóviles sacamos los siguientes datos:

En el año 1995, en el 25% de los países se redujo la cantidad compradores de autos nuevos. En el 30% de los países subió el costo de los autos nuevos. Y en el 80% de los países en los cuales subió el costo, bajó la cantidad de compradores. Es decir:

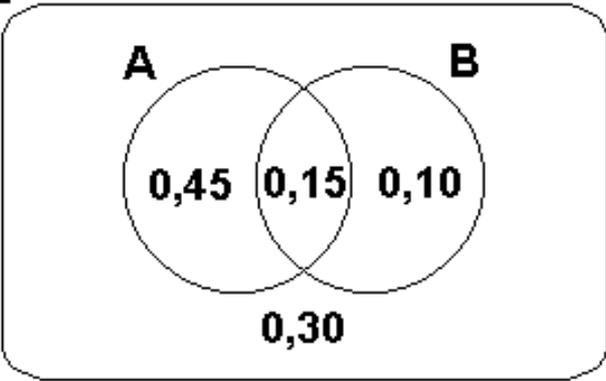
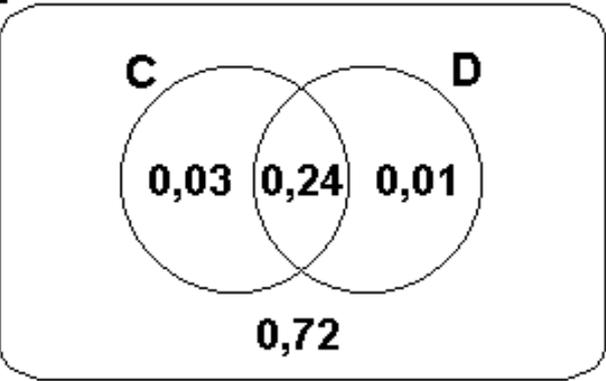
$$P(D/C) = 0,8$$

$$P(D) = 0,25$$

$$P(C) = 0,3$$

Vemos que $P(D/C) \neq P(D)$ por lo tanto los sucesos C y D no son independientes, por lo tanto tampoco se cumplen las otras dos definiciones y entonces $P(C/D) \neq P(C)$ y también $P(C \cap D) \neq P(C) \cdot P(D)$

A continuación hagamos los diagramas de Venn de los dos ejemplos dados:

<p>E</p>  <p>A B</p> <p>0,45 0,15 0,10</p> <p>0,30</p>	<p>E</p>  <p>C D</p> <p>0,03 0,24 0,01</p> <p>0,72</p>
<p>Independientes (se cumplen las definiciones)</p>	<p>No independientes (no se cumplen las definiciones)</p>

Casos especiales de dependencia

• Sucesos disjuntos:

Si los sucesos son disjuntos, el hecho de que ocurra uno implica que el otro no ocurre. Es decir, en el caso de que sean disjuntos, el hecho que un suceso ocurra no solamente afecta la probabilidad de que el otro ocurra, sino que además la hace directamente cero. Por lo tanto los sucesos son fuertemente dependientes.

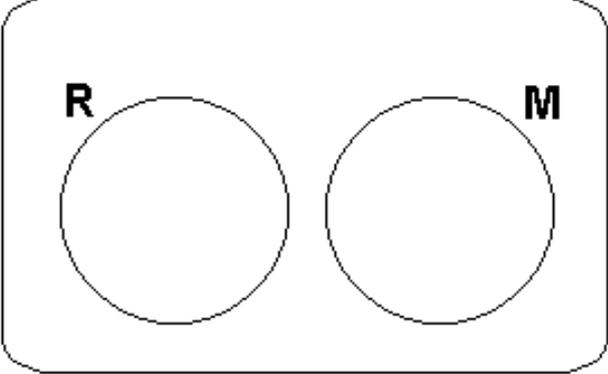
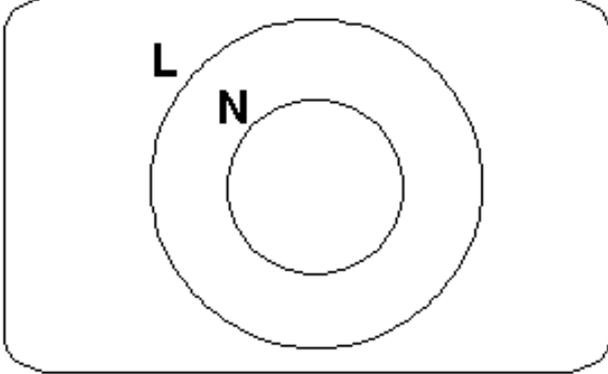
Si el suceso R es que una persona sea rubia y el suceso M es que sea morocha, $R \cap M = \emptyset$, y por lo tanto si se sabe que una persona es rubia la probabilidad de que sea morocha es cero y también si se sabe que una persona es morocha, la probabilidad de que sea rubia es cero. Vemos que por tratarse de sucesos disjuntos, el hecho de que ocurra uno hace que la probabilidad no solamente sea afectada sino que además la hace valer cero.

• Un suceso incluido en otro:

Si un suceso está incluido en otro, al ocurrir el de “adentro” necesariamente ocurre también el de “afuera”. Es decir, el hecho de que haya ocurrido el de “adentro” modifica la probabilidad de que ocurra el de “afuera”, y de hecho la hace uno.

Si el suceso N es haya nubes un determinado día haya nubes y el suceso L es que llueva, notamos que $L \subset N$. El hecho de saber que un día llovió hace que la probabilidad de que haya habido nubes sea 1, con lo cual el hecho de saber que ocurrió L afecta la probabilidad de N. Y también el hecho de saber que hubo nubes no necesariamente implicará que llueva, pero en general afectará la probabilidad de que llueva, porque recordemos que aceptar que “hay nubes” implica meterse en un espacio muestral en el cual “hay nubes”, y por lo tanto todas las probabilidades se modifican porque deben estar referidas al nuevo espacio muestral.

Visualicemos estos ejemplos mediante diagramas de Venn:

	
<p style="text-align: center;">No independientes Los sucesos disjuntos no pueden ser independientes.</p>	<p style="text-align: center;">No independientes Si un suceso está incluido en otro no pueden ser independientes</p>

Independencia de los complementos

Dados dos sucesos A, B:

$A, B \text{ indep.} \Leftrightarrow A, B^c \text{ indep.} \Leftrightarrow A^c, B \text{ indep.} \Leftrightarrow A^c, B^c \text{ indep.}$

La justificación es simple, si el hecho de que ocurra A no afecta la probabilidad de B, entonces tampoco afecta la probabilidad de que **no** ocurra B.

Por ejemplo si se sabe que los sucesos:

A: Argentina le gana hoy a Brasil en el partido de fútbol

B: Esta noche hay luna llena

son independientes, y se tiene el suceso:

X: Esta noche no hay luna llena

¿Son A y X independientes? Sí, porque $X = B^c$, y si A y B son independientes, A y B^c también lo son. Dicho de otro modo, si el hecho de que gane Argentina no afecta la probabilidad de que haya luna llena, tampoco afecta la probabilidad de que no haya luna llena. Y tampoco por ejemplo, si la probabilidad de que haya luna llena no afecta la probabilidad de que gane Argentina, tampoco afecta la probabilidad de que no gane Argentina.

Problemas típicos

1) Indique qué puede afirmar acerca de la independencia de los siguientes pares de sucesos:

- a) Que al tirar una moneda y un dado salga cara en la moneda y 3 en el dado.
- b) Que la clase sea buena y que los alumnos entiendan.
- c) Que una lata de arvejas pese más de 200 g y que contenga más de 300 arvejas.
- d) Que llueva y que suene el teléfono en los próximos 5 minutos.
- e) Que llueva y que haya nubes
- f) Que un número sea par y que ese mismo número sea impar
- g) Que al tirar una moneda y un dado salga cara en la moneda y NO salga 3 en el dado.

Haga las aclaraciones que considere necesarias.

Resolución:

- a) Podemos suponer que son independientes, porque no parece que si ocurre una cosa se vea afectada la probabilidad de que ocurra la otra.
- b) Podemos suponer que no son independientes, porque la probabilidad de que los alumnos entiendan si la clase fue buena debe ser mayor que si no lo fue, y visto de otro modo, si los alumnos entendieron, la probabilidad de que la clase haya sido buena debe ser mayor que si los alumnos no entendieron.
- c) Podemos suponer que no son independientes, porque hay una relación entre el peso de la lata y la cantidad de arvejas que contiene, y como los sucesos "la lata pesa más de 200 g" y "la lata contiene más de 300 arvejas" son condiciones impuestas sobre esas cantidades relacionadas, no pueden ser independientes.
- d) Podemos suponer que son independientes. En principio no hay ninguna relación entre una cosa y la otra. Pero si tuviésemos más información (por ejemplo, que una tía siempre nos llama para recordarnos que cerremos las ventanas porque se ha largado a llover) nuestra respuesta podría ser diferente, porque en ese caso el hecho de que ha comenzado a llover incrementa la probabilidad de que suene el teléfono en los próximos 5 minutos porque puede ser la tía avisándonos que está lloviendo.
- e) No son independientes, porque uno está incluido en otro.

f) No son independientes, porque son disjuntos.

g) Los suponemos independientes por las mismas razones que en a), o también porque el suceso del dado es el complemento de un suceso que era independiente del de la moneda, entonces también es independiente.

2) Determinar si los sucesos A y B son independientes, de acuerdo a los siguientes datos:

a) $P(A) = 0,3$; $P(B) = 0,2$; $P(A \cap B) = 0,05$

b) $P(A \cap B^c) = 0,1$; $P(A \cap B) = 0,2$; $P(A/B) = 0,3$

Resolución:

a) $P(A) \cdot P(B) = 0,3 \cdot 0,2 = 0,06 \neq 0,05 = P(A \cap B)$, por lo tanto no son independientes

b) $P(A \cap B^c) + P(A \cap B) = P(A) = 0,3 = P(A/B)$, por lo tanto son independientes

3) Si la probabilidad de que hoy llueva es 0.2 y la probabilidad de que hoy se me acabe la tinta de la lapicera es 0.6, calcule la probabilidad de que:

a) llueva y se me acabe la tinta

b) llueva y no se me acabe la tinta

c) no llueva y no se me acabe la tinta

Aclare qué suposiciones debe hacer.

Resolución:

Debemos suponer que el suceso de que hoy llueva y el de que se me acabe la tinta son independientes (si no, no se podría resolver). Nos dicen que la probabilidad de que llueva es 0.2, por lo cual la probabilidad de que no llueva es 0.8. Además la probabilidad de que se acabe la tinta es 0.6, por lo cual la probabilidad de que no se acabe la tinta es 0.4.

Resolvemos:

a) Sabemos que cuando dos sucesos son independientes, la probabilidad de que ocurran simultáneamente es el producto de las probabilidades de que ocurran individualmente.

Es decir, los sucesos A y B son independientes $\Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$

Si tomamos A: "que llueva" y B: "que se me acabe la tinta" entonces:

$$P(A \cap B) = P(A) \cdot P(B) = 0.2 \cdot 0.6 = 0.12$$

b) Si A y B son independientes, entonces A y B^c también lo son. Entonces vale:

$$P(A \cap B^c) = P(A) \cdot P(B^c) = 0.2 \cdot 0.4 = 0.08$$

c) Si A y B son independientes, entonces A^c y B^c también lo son. Entonces vale:

$$P(A^c \cap B^c) = P(A^c) \cdot P(B^c) = 0.8 \cdot 0.4 = 0.32$$

4) Se tiran 2 dados honestos. Calcule la probabilidad de que:

a) No salga ningún 1

b) No salga ningún número impar.

Resolución:

a) Consideraremos a los dados independientes. Y entonces tomamos los sucesos:

A: que no salga un 1 en el primer dado.

B: que no salga un 1 en el segundo dado.

Y queda:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{5}{6} \cdot \frac{5}{6} = 0.694$$

También lo podríamos haber pensado de acuerdo a lo que vimos cuando estudiamos multiplicación de probabilidades. Tomando los mismos sucesos A y B, lo que estamos buscando es $P(A \cap B)$, lo cual según vimos se puede escribir como $P(A) \cdot P(B/A)$. En este caso particular, por considerarlos independientes, $P(B/A)$ termina siendo $P(B)$, y entonces llegamos al mismo resultado que con el otro planteo es decir $P(A) \cdot P(B) = 0.694$

b) Nuevamente los consideramos independientes. Y tomamos los sucesos:

A: que no salga ningún número impar en el primer dado.

B: que no salga ningún número impar en el segundo dado.

Y queda:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{3}{6} \cdot \frac{3}{6} = 0.25$$

Aquí también podríamos hacer el mismo razonamiento que antes.

5) La probabilidad de acertarle a un blanco en cada disparo es de 0.6. ¿Cuál es la probabilidad de que, efectuando 5 disparos, se acierte el primero, se falle el segundo, se acierten el tercero y el cuarto, y se falle el quinto?

Resolución:

Si aplicamos el mismo enfoque que en los anteriores, asumiremos que los 5 intentos son independientes y haremos:

A: acertar el primero

B: fallar el segundo

C: acertar el tercero

D: acertar el cuarto

F: fallar el quinto

$$P(A \cap B \cap C \cap D \cap F) = P(A) \cdot P(B) \cdot P(C) \cdot P(D) \cdot P(F) = 0.6 \cdot 0.4 \cdot 0.6 \cdot 0.6 \cdot 0.4 = 0.03456$$

Lo cual es correcto. Podríamos haberlo pensado con multiplicación de probabilidades, con lo cual el resultado habría sido $P(A) \cdot P(B/A) \cdot P(C/B \cap A) \cdot \dots$ y las condiciones habrían desaparecido porque los sucesos son independientes, y llegaríamos al mismo resultado que antes.

También podríamos hacer:

A: probabilidad de acertar un disparo

$$P(A \cap A^c \cap A \cap A \cap A^c) = P(A) \cdot P(A^c) \cdot P(A) \cdot P(A) \cdot P(A^c) = 0.6 \cdot 0.4 \cdot 0.6 \cdot 0.6 \cdot 0.4 = 0.03456$$

Y obtenemos el mismo resultado. Esto se debe a que luego de cada intento, la probabilidad de acertar sigue siendo la misma (se mantiene constante) y cada sucesivo disparo se lleva a cabo en las mismas condiciones que el primero.

6) Se tiene una cierta moneda cargada, para la cual la probabilidad de sacar cara es 0.7. Si un experimento consiste en tirar dicha moneda 2 veces, calcule la probabilidad de:

- a) sacar primero cara y después ceca
- b) sacar primero ceca y después cara
- c) sacar una cara y una ceca

Resolución:

a) A: sacar cara al tirar la moneda

$$P(A) = 0.7$$

Como vimos en el ejemplo anterior, consideramos los dos intentos independientes y hacemos:

$$P(\text{sacar cara en la primera y ceca en la segunda}) = P(A \cap A^c) = 0.7 \cdot 0.3 = 0.21$$

b) Bajo las mismas condiciones:

$$P(\text{sacar ceca en la primera y cara en la segunda}) = P(A^c \cap A) = 0.3 \cdot 0.7 = 0.21$$

Vemos que nuevamente no importa el orden.

c) "Sacar una cara y una ceca" es equivalente a "Sacar cara y después ceca, o sacar ceca y después cara". Entonces si el suceso A es "sacar cara", la probabilidad es:

$$P(\text{sacar una cara y una ceca}) = P((\text{sacar cara y después ceca}) \cup (\text{sacar ceca y después cara})) = P((A \cap A^c) \cup (A^c \cap A))$$

Aplicamos la fórmula para la probabilidad de la unión y obtenemos:

$$P((A \cap A^c) \cup (A^c \cap A)) = P(A \cap A^c) + P(A^c \cap A) - P((A \cap A^c) \cap (A^c \cap A))$$

$P(A \cap A^c)$ y $P(A^c \cap A)$ ya estaban calculadas antes

$$P((A \cap A^c) \cap (A^c \cap A)) = P(A \cap A^c \cap A^c \cap A) = P(A \cap A^c) = P(\emptyset) = 0$$

Lo cual es lógico porque no puede salir cara y no salir cara al mismo tiempo.

Entonces queda:

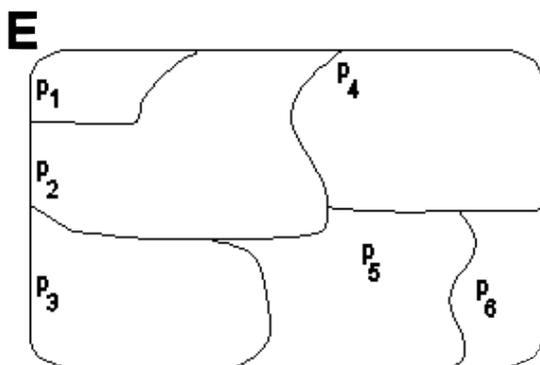
$$P((A \cap A^c) \cup (A^c \cap A)) = P(A \cap A^c) + P(A^c \cap A) = 0.21 + 0.21 = 0.42$$

Vemos que no importa el orden en el sentido de que todas las formas de ordenar tienen la misma probabilidad, pero si queremos tomar la probabilidad de que ocurra, y ocurra en cualquier orden, la probabilidad será, lógicamente, mayor, ya que será la unión de todos los órdenes posibles en los que puede ocurrir.

Probabilidad total

Comencemos por recordar lo que es una partición. Una partición de un conjunto es una forma de dividirlo en una determinada cantidad de subconjuntos denominados partes, tales que esas partes son todas disjuntas, y a la vez la unión de todas ellas forman el conjunto original.

Por ejemplo ilustremos una posible partición de un determinado conjunto E:



Vemos en el dibujo que se cumplen las dos condiciones que enunciamos sobre las partes:

$$1) E = \bigcup_{i=1}^n p_i$$

(la unión de las partes es el conjunto)

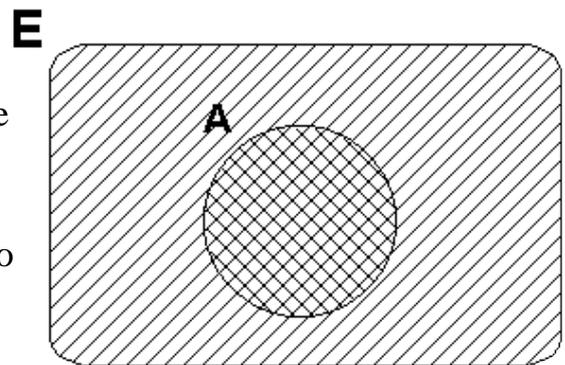
$$2) p_i \cap p_j = \emptyset \text{ para } i \neq j$$

(todas las partes son disjuntas entre sí)

Por otro lado, recordemos que si un determinado conjunto A está incluido en otro conjunto E, entonces por propiedades de conjuntos sabemos que $A \cap E = A$

Usando dicho resultado, podemos decir que si el conjunto E es el espacio muestral de un experimento y A es un suceso (o sea un subconjunto de ese espacio muestral), entonces:

$$P(A) = P(A \cap E) \text{ (porque como } A \subset E, A \cap E = A)$$



Luego podemos, por ejemplo, crear una partición del conjunto E, subdividiéndolo en n partes p_i , y luego por la primera propiedad de las particiones, $E = \bigcup_{i=1}^n p_i$

Es decir, podemos escribir E como la unionatoria de las partes, y entonces:

$$P(A \cap E) = P\left(A \cap \bigcup_{i=1}^n p_i\right) = P(A \cap (p_1 \cup p_2 \cup \dots \cup p_n))$$

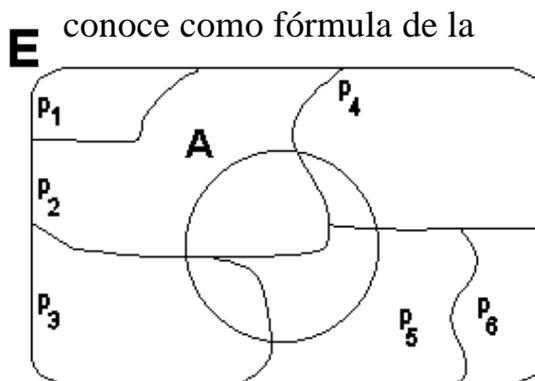
Luego se puede aplicar la propiedad distributiva de conjuntos, y se obtiene:

$$P(A \cap (p_1 \cup p_2 \cup \dots \cup p_n)) = P((A \cap p_1) \cup (A \cap p_2) \cup \dots \cup (A \cap p_n))$$

Notemos ahora que como las p_i son disjuntas, entonces los $(A \cap p_i)$ también son todos disjuntos. En consecuencia, por el tercer axioma podemos escribir la probabilidad de esa suma como la suma de las probabilidades, y nos queda:

$$P((A \cap p_1) \cup (A \cap p_2) \cup \dots \cup (A \cap p_n)) = P(A \cap p_1) + P(A \cap p_2) + \dots + P(A \cap p_n) = \\ = \sum_{i=1}^n P(A \cap p_i)$$

Entonces, en resumen, llegamos a lo que se conoce como fórmula de la



probabilidad total:

$$P(A) = \sum_{i=1}^n P(A \cap p_i)$$

Es decir, la probabilidad de A es la suma de las probabilidades de las intersecciones de A con cada parte del espacio muestral.

Esto es útil porque a menudo se quiere calcular la probabilidad de un determinado suceso compuesto por diversos resultados y resulta muy fácil y práctico (y a veces casi obligatorio) encontrar una partición del espacio muestral y calcular la probabilidad del suceso mediante la fórmula de la probabilidad total.

Otro resultado que es útil y constituye un caso particular de probabilidad total es la de un suceso y su complemento. Dado un espacio muestral E y un suceso cualquiera D, como se estudió al comienzo de este capítulo D y D^c forman una partición de E porque $D \cup D^c = E$ y $D \cap D^c = \emptyset$.

D y D^c son entonces las p_i , y podemos calcular la probabilidad de otro suceso A con la probabilidad total:

$$P(A) = \sum_{i=1}^n P(A \cap p_i) = P(A \cap D) + P(A \cap D^c)$$

Vemos ahí justificada de otra forma la expresión que utilizamos antes para resolver problemas. En el fondo estábamos usando probabilidad total.

Pero volvamos a la fórmula de la probabilidad total. Si hacemos un paso más y le aplicamos la definición de probabilidad condicional a $P(A \cap p_i)$, llegamos a una expresión alternativa, que por lo general resulta más práctica y se usa en la mayoría de los casos:

$$P(A) = \sum_{i=1}^n P(A \cap p_i) = \sum_{i=1}^n P(A | p_i) \cdot P(p_i)$$

Problemas típicos

1) En una determinada ciudad, la probabilidad de que una persona elegida al azar sea mujer y tenga ojos azules es 0.1, y la probabilidad de que una persona elegida al azar sea hombre y tenga ojos azules es 0.15. ¿Cuál es la probabilidad de que una persona elegida al azar tenga ojos azules?

Resolución:

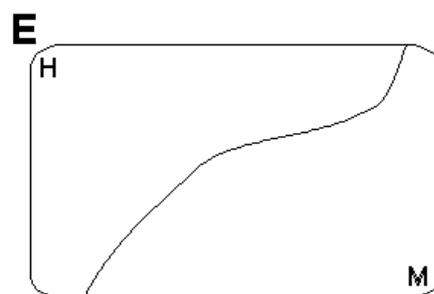
El experimento consiste en tomar una persona al azar, y registrar su sexo y el color de sus ojos. Definimos 3 sucesos:

M: que la persona elegida al azar sea mujer.

H: que la persona elegida al azar sea hombre.

A: que la persona elegida al azar tenga ojos azules.

Como una persona es hombre o es mujer y no hay otras posibilidades, entonces $P(M) + P(H) = 1$. Como además no se puede ser hombre y mujer al mismo tiempo, los sucesos M y H son disjuntos. Es decir, M y H constituyen una partición del espacio muestral E.



Nos piden: $P(A)$

Como $A \subset E$ entonces $A \cap E = A$. Por lo tanto:

$$P(A) = P(A \cap E)$$

Como M y H constituyen una partición de E, entonces:

$$P(A \cap E) = P(A \cap (M \cup H))$$

Distribuimos la intersección y queda:

$$P(A \cap (M \cup H)) = P((A \cap M) \cup (A \cap H))$$

Como M y H son disjuntos, entonces $(A \cap M)$ y $(A \cap H)$ son disjuntos. (Una persona no puede ser hombre y mujer al mismo tiempo, por lo tanto tampoco puede (tener ojos azules y ser mujer) y (tener ojos azules y ser hombre) al mismo tiempo. Queda:

$$P((A \cap M) \cup (A \cap H)) = P(A \cap M) + P(A \cap H) = 0.1 + 0.15 = 0.25$$

2) Una empresa que fabrica lamparitas tiene 2 plantas, la A y la B. Cada lamparita fabricada por A tiene probabilidad 0.01 de ser defectuosa. Cada lamparita fabricada por B tiene probabilidad 0.02 de ser defectuosa. Si las plantas A y B producen el 60% y el 40% de las unidades respectivamente, ¿cuál es la probabilidad de que una lamparita fabricada por la empresa sea defectuosa?

Resolución:

Siguiendo el mismo análisis del ejercicio anterior, el experimento consiste en tomar una lamparita, y ver quién la fabricó, y si es defectuosa. Tomamos los siguientes sucesos:

A: que la lamparita haya sido fabricada por la planta A

B: que la lamparita haya sido fabricada por la planta B

D: que la lamparita sea defectuosa

Observamos que A y B son una partición de E, porque $A \cup B = E$ (la lamparita obligatoriamente fue fabricada por alguna de las 2 plantas) y $A \cap B = \emptyset$ (la lamparita no puede haber sido fabricada por las 2 plantas).

Nos piden $P(D)$. Como en el ejercicio anterior vimos el desarrollo paso por paso, ahora aplicaremos directamente la fórmula de la probabilidad total:

$$P(D) = \sum_{i=1}^n P(D \cap p_i) = \sum_{i=1}^n P(D|p_i) \cdot P(p_i)$$

Donde $n = 2$, $P_1 = A$, $P_2 = B$.

Usamos la segunda expresión y queda:

$$P(D) = \sum_{i=1}^n P(D|P_i) \cdot P(P_i) = P(D|A) \cdot P(A) + P(D|B) \cdot P(B)$$

Observemos que si nos dicen que la probabilidad de que una lamparita fabricada por A sea defectuosa es 0.01, nos están diciendo $P(D / A) = 0.01$. Entonces:

Otros problemas

La probabilidad total es una herramienta muy utilizada en muchos temas de probabilidad y estadística, por lo que las aplicaciones más importantes aparecerán en los próximos capítulos.

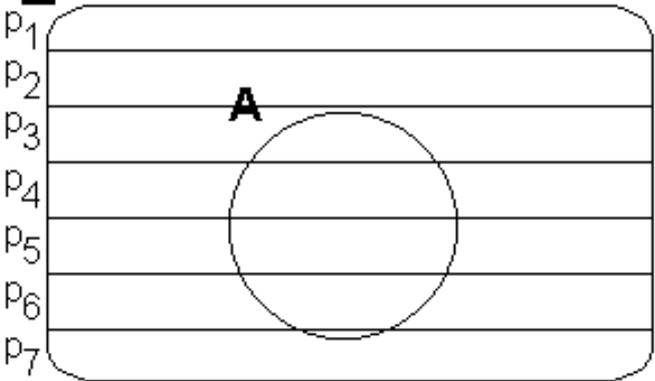
Regla de Bayes

Consideremos un modelo como el que planteamos al estudiar la probabilidad total, en el cual el espacio muestral estaba particionado y se quería calcular la probabilidad de un suceso A contenido en ese espacio muestral.

Supongamos ahora que lo que se desea no es la probabilidad del suceso A sino la probabilidad de una de las partes, sabiendo que ocurrió A.

El lector podrá advertir que esto está íntimamente relacionado con lo que se dijo al estudiar la probabilidad condicional: que cuando se aplica una condición, el nuevo espacio muestral pasa a ser el suceso en el cual se cumple esa condición, y entonces las probabilidades cambian porque ahora están referidas a un nuevo espacio muestral (si esto no se entiende inmediatamente recomendamos repasar las secciones 1.4, 1.5 y 1.6)

Dijimos entonces que el espacio muestral E estaba particionado, y que se sabe que ocurrió A, y entonces se desea calcular la probabilidad de cada parte (es decir, calcular las nuevas probabilidades, referidas al espacio muestral A).

<p>E</p> 	<p>A</p> 
<p style="text-align: center;">a priori</p> <p>conocemos las probabilidades originales de las partes, o sea las $P(p_i)$</p>	<p style="text-align: center;">a posteriori</p> <p>conocemos las probabilidades de las partes sabiendo que ocurrió A, o sea las $P(p_i/A)$</p>

Si queremos calcular la probabilidad de la parte p_i , sabiendo que ocurrió A, planteamos:

$$P(p_i/A) = \frac{P(p_i \cap A)}{P(A)}$$

En el denominador usamos la fórmula de la probabilidad total, y nos queda:

$$P(p_i/A) = \frac{P(p_i \cap A)}{\sum_{i=1}^n P(p_i \cap A)}$$

A continuación damos vuelta las dos intersecciones y aplicamos la definición de probabilidad condicional, y queda:

$$\frac{P(p_i \cap A)}{\sum_{i=1}^n P(p_i \cap A)} = \frac{P(A \cap p_i)}{\sum_{i=1}^n P(A \cap p_i)} = \frac{P(A/p_i) P(p_i)}{\sum_{i=1}^n P(A/p_i) P(p_i)}$$

En conclusión:

$$P(p_i/A) = \frac{P(A/p_i) P(p_i)}{\sum_{i=1}^n P(A/p_i) P(p_i)}$$

Lo cual se conoce como regla de Bayes ó fórmula de Bayes.

Observemos que se tienen como dato las probabilidades originales de las partes y la probabilidad de que ocurra A dentro de cada parte, y lo que se obtiene es la probabilidad de que ocurra una determinada parte sabiéndose que ocurrió A.

Ejemplo

En un determinado grupo de gente hay personas rubias, morochas y pelirrojas. El 60% de la gente es morocha, el 30% rubia y el 10% pelirroja. El 50% de los rubios tiene ojos claros, el 40% de los pelirrojos tiene ojos claros y el 25% de los morochos tiene ojos claros. Si una persona elegida al azar tiene ojos claros, ¿cuál es la probabilidad de que sea rubia?

Aplicamos la regla de Bayes:

$$P(p_i/A) = \frac{P(A/p_i) P(p_i)}{\sum_{i=1}^n P(A/p_i) P(p_i)} = \frac{0,5 \cdot 0,3}{0,25 \cdot 0,6 + 0,5 \cdot 0,3 + 0,4 \cdot 0,1} = 0,441$$



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 4 de mayo de 2004

CAPÍTULO II

Variable Aleatoria

Si alguien nos dice que hay un cuadrado cuyo lado mide a , y nos pregunta cuál es el área del cuadrado, responderemos que el área es a^2 . Y podemos responder esa pregunta sin saber cuánto vale a . Es decir, podemos abstraernos de la longitud del lado, y contestar cuál es la superficie.

Muchas veces necesitamos trabajar con magnitudes sin tener en cuenta los valores. Esto puede ser por distintas razones. Puede ser que no conozcamos los valores. También puede ser que haya muchos valores posibles distintos y querramos trabajar sin tener en cuenta cuál de todos consideraremos al final.

En conclusión, cuando necesitamos trabajar con números abstrayéndonos de los valores, usamos VARIABLES. La ecuación de una recta es $y=ax+b$. Y eso se cumple para todos los puntos de la recta. Entonces en vez de escribirla para cada punto, la dejamos expresada usando variables.

Vamos a llamar variable aleatoria a una variable cuyo valor sería el resultado de un determinado experimento, si lo hiciéramos. Por ejemplo, si el experimento consiste en arrojar un dado, podemos definir la variable aleatoria X cuyo valor será el número que salga en el dado. El conjunto de valores posibles de X es el espacio muestral. Y en general nos interesará cuál es la probabilidad de que X asuma cada valor.

Vemos que la probabilidad de que X asuma un determinado valor, es la probabilidad de que el experimento arroje un determinado resultado, con lo cual observamos que esto en esencia es lo mismo que vimos en el primer capítulo, pero

con otro enfoque.

Usaremos variables porque nos permiten operar y mostrar determinadas conclusiones.

Para el caso del dado, podemos escribir "la probabilidad de que al tirar el dado salga un número mayor que 3" simplemente como $P(X > 3)$, habiendo antes definido X como el número que saldría si tiráramos el dado.

Para designar a las variables aleatorias se utilizan letras mayúsculas. Para designar a uno de sus valores posibles, se usan las letras minúsculas. Por ejemplo, si X es la variable aleatoria asociada a lo que sale al tirar un dado honesto, podemos decir que $P(X = x) = 1/6, \forall x$.

En el capítulo anterior vimos que el espacio muestral es el conjunto de resultados del experimento aleatorio. Y dado el sinnúmero de experimentos posibles, los resultados podían ser cosas tan diversas como {cara ; ceca}, {1, 2, 3, 4, 5, 6}, {ganar la lotería, no ganar la lotería}, {"soltero, "casado", "viudo", "divorciado"}. Es decir, el resultado de un experimento puede ser un número, un valor booleano(si/no), un texto, etc. Entonces una variable aleatoria puede ser numérica, booleana, etc.

Pero como con los números podemos medir magnitudes y hacer operaciones, por lo general podemos extraer de ellos mayor cantidad de información que con otras cosas. Por eso concentraremos nuestro estudio en los experimentos cuyo resultado es un número. Y entonces trabajaremos con variables aleatorias numéricas. Es decir, variables cuyos valores posibles son números.

A raíz de ello, a veces interpretaremos resultados que no son números, por ejemplo, el estado civil de una persona observada, como números, por ejemplo estableciendo una codificación numérica: 1=soltero, 2=casado, 3=viudo, etc.

Ejemplo

Se tiene el experimento "tirar un dado y considerar el número que sale"

El espacio muestral es $E = \{1, 2, 3, 4, 5, 6\}$ (*)

Definiremos una variable aleatoria:

X : el número que sale al tirar el dado

Ahora usaremos esa variable aleatoria para calcular la probabilidad de que salga un número mayor que 3. Es decir:

$$P(X > 3)$$

Observemos que " $X > 3$ " es un suceso. Ahora lo vamos a reemplazar por el suceso equivalente " $X=4 \vee X=5 \vee X=6$ ".

$$P(X > 3) = P(X=4 \vee X=5 \vee X=6)$$

Como los sucesos " $X=4$ ", " $X=5$ " y " $X=6$ " son disjuntos, podemos sumar sus probabilidades:

$$P(X=4 \vee X=5 \vee X=6) = P(X=4) + P(X=5) + P(X=6)$$

Y ahora reemplazamos por las probabilidades que ya son conocidas:

$$P(X=4) + P(X=5) + P(X=6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$

Con lo cual $P(X > 3) = \frac{1}{2}$.

(*) existe una discusión filosófica acerca de si los resultados de dicho experimento son realmente los números del 1 al 6 o bien "sale 1", "sale 2", "sale 3", etc. y los valores posibles de la variable aleatoria son fruto de interpretar esos resultados proposicionales ("sale 3") como resultados numéricos ("3"). Adoptaremos la convención de considerar que los resultados del experimento son directamente los números, cuando, como en el presente ejemplo, los resultados posibles tengan interpretación numérica inmediata.

VARIABLES ALEATORIAS DISCRETAS Y CONTINUAS

Comparemos ahora el ejemplo del dado con este otro: haremos el experimento de elegir una naranja al azar en una verdulería, y llamaremos Y al peso de la naranja elegida. Si pensamos en los valores posibles que puede tomar la variable aleatoria Y , veremos que no solamente son infinitos sino que además dado un valor posible no hay un "siguiente" porque entre cualquier valor y aquel al que consideráramos su "siguiente" hay infinitos valores posibles. La variable aleatoria X es discreta. La variable aleatoria Y es continua.

En principio definiremos las variables aleatorias discretas y continuas así:

- Variable aleatoria discreta: aquella tal que la cantidad de valores posibles que puede tomar es finita, o infinita pero numerable. En otras palabras, aquella cuyos valores posibles son todos puntos aislados del conjunto de valores posibles. Dicho incluso de una tercera forma: aquella tal que si tomamos dos cualesquiera de sus valores posibles, hay entre ellos una cantidad finita de valores posibles.
- Variable aleatoria continua: aquella que no es discreta, es decir, aquella tal que la cantidad de valores posibles es infinita y no numerable.

¿A qué nos referimos con infinito numerable y no numerable?

Por ejemplo, el conjunto de los números naturales tiene una cantidad finita pero numerable de elementos, porque sus elementos se pueden enumerar.

En cambio, el conjunto de los números reales tiene una cantidad infinita no numerable de elementos, porque sus elementos no se pueden enumerar.

Entonces una variable aleatoria es discreta cuando se pueden enumerar sus valores posibles (aunque sean infinitos) y es continua cuando no se pueden enumerar. Dicho de otra forma, el rango (conjunto de valores posibles) de una variable aleatoria discreta es la unión de puntos aislados (en una recta, en un plano, etc.), mientras que el rango de una variable aleatoria continua es uno o varios segmentos de recta, una superficie en un plano, etc.

De la definición de variable aleatoria continua podríamos inferir que, como cada valor posible es un punto en un continuo, es decir, un resultado posible entre una cantidad infinita y ni siquiera numerable de resultados posibles, entonces la probabilidad de que ocurra ese resultado posible es "cero". Ese cero es comparable a la longitud de un diferencial de longitud o a la superficie de un diferencial de área. Es decir, no es que sea cero, porque si no, no sería un resultado posible. Es una probabilidad de orden diferencial (tiende a cero), por lo cual a los fines prácticos consideramos que vale cero.

La forma de proceder con las variables aleatorias discretas y continuas no es siempre la misma, por lo cual en adelante haremos distinciones entre ellas. Sin embargo, en muchos casos las definiciones y métodos que utilizaremos para ambos tipos de variables son análogos.

Problemas Típicos

1) Indique para cada una de las siguientes variables aleatorias si son discretas o continuas. Haga las aclaraciones que considere necesarias.

- a) El número que sale al tirar un dado.**
- b) La cantidad de caras que salen al tirar 5 monedas.**
- c) La cantidad de accidentes por mes**
- d) Peso de una naranja.**
- e) Diámetro de una arandela.**
- f) El país donde nació una persona.**
- g) La edad de una persona.**

Resolución:

a) Discreta. La cantidad de resultados es finita.

b) Idem

c) Discreta. Aunque la cantidad de resultados es infinita, porque no hay un valor máximo posible, es numerable, porque los resultados se pueden enumerar. Otra forma de ver que es discreta: todos los resultados son puntos aislados.

d) Continua. La cantidad de resultados es infinita y no numerable (no podemos enumerar todos los resultados). Otra forma de ver que es continua: los resultados no son puntos aislados, sino que forman un continuo (por ejemplo, un segmento de recta).

e) Idem.

f) Discreta. La cantidad de resultados es finita. Observemos que las variables que no son numéricas por lo general son discretas.

g) Puede ser discreta o continua. Si tomamos la edad como la cantidad entera de años que ha vivido la persona, entonces es discreta. Si tomamos la edad como un número real de años que ha vivido la persona (ejemplo: 5,37 años) entonces es continua.

Distribución de probabilidad

Una variable aleatoria tal que todos sus valores posibles son equiprobables es un caso muy particular. En general, cada uno de los valores posibles puede tener distinta probabilidad. Por eso nos interesa estudiar cómo se distribuyen las probabilidades en los distintos valores posibles de la variable.

Al conjunto de valores posibles, y la relación entre ellos y sus respectivas probabilidades, se lo conoce como distribución de probabilidad.

Notemos que:

- 1) la probabilidad de un determinado valor no puede ser menor que cero.
- 2) la suma de las probabilidades de todos los valores da 1, porque al hacer el experimento siempre sale uno de los resultados posibles.

La distribución de probabilidad se puede expresar de diversas formas.

Generalmente se usa la función de densidad de probabilidad.

Función de densidad de probabilidad

Esta función le asigna a cada valor posible de la variable aleatoria un número real que consiste en la probabilidad de que ocurra, y por supuesto debe cumplir con las 2 condiciones que enunciamos antes:

- a) no puede ser negativa en ningún punto
- b) la suma de las probabilidades de todos los valores da 1.

Puede pensarse que la condición "a" es insuficiente, porque la probabilidad no solamente no puede ser menor que cero, sino tampoco mayor que uno. Pero agregar esa condición sería redundante, porque la condición "b" garantiza que eso no puede ocurrir, ya que si la probabilidad para un valor fuera mayor que 1, como ninguna probabilidad puede ser negativa entonces la suma daría necesariamente mayor a 1.

Si X es discreta:

$P_X(x)$ es una función que a cada valor posible le asigna su probabilidad.

$P_X(x)$ es una función de densidad de probabilidad discreta si y solo si cumple con:

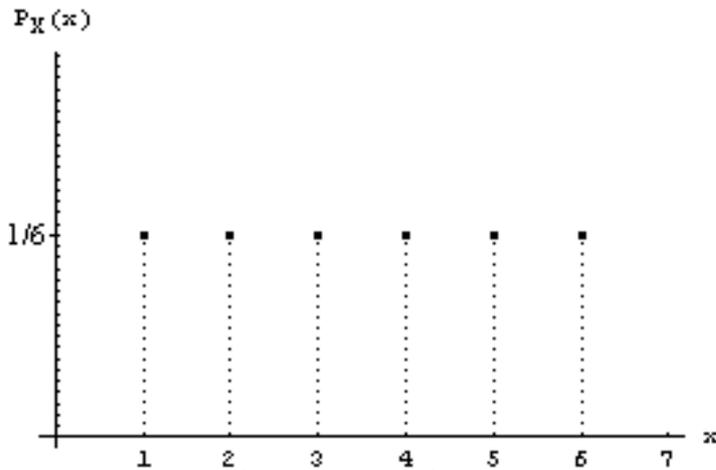
$$1) P_X(x) \geq 0 \quad \forall x$$

$$\sum P_X(x) = 1$$

$$2) \quad x$$

Ejemplo:

X: el número que sale al tirar un dado honesto:



$$P_X(x) = P(X = x) = \left\{ \begin{array}{ll} 1/6 & x = 1 \\ 1/6 & x = 2 \\ 1/6 & x = 3 \\ 1/6 & x = 4 \\ 1/6 & x = 5 \\ 1/6 & x = 6 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

O bien su forma abreviada:

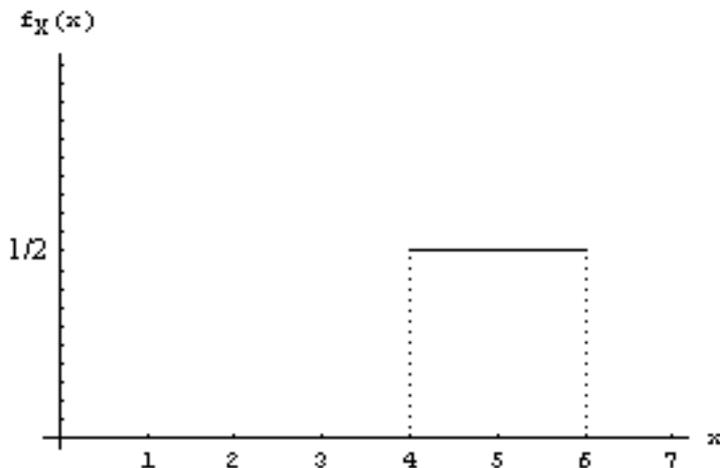
$$P_X(x) = \left\{ \begin{array}{ll} 1/6 & x = 1, 2, 3, 4, 5, 6 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

Si X es continua:

Habíamos dicho que la probabilidad de que una variable aleatoria continua asumiera un determinado valor es cero. Entonces trabajaremos con intervalos:

$f_X(x)$ es una función que integrada entre a y b nos da la probabilidad de que la variable aleatoria X asuma un valor entre a y b.

$f_X(x)$ es una función de densidad de probabilidad continua si y solo si cumple con:



1) $f_X(x) \geq 0 \forall x$

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1$$

2) $-\infty$

Ejemplo:

X: el resultado de elegir un número real al azar entre 4 y 6:

$$f_X(x) = \begin{cases} 0,5 & 4 < x < 6 \\ 0 & \forall \text{ otro } x \end{cases}$$

Luego la probabilidad de que X caiga en un determinado intervalo es el area bajo la curva de f_X en ese intervalo.

Cálculo de probabilidades con la función de densidad

Si X es discreta:

$P(X = x) = P_X(x)$	porque la función $P_X(x)$ nos da justamente la probabilidad de que X asuma el valor x.
$P(X \leq x) = \sum_{-\infty} P_X(x)$	porque la probabilidad de que $X \leq x$ es la probabilidad de que X valga x o cualquier valor menor que x.
$P(X < x) = \left(\sum_{-\infty} P_X(x) \right) - P(X = x)$	es igual que la anterior pero le restamos $P(X = x)$ porque estamos pidiendo que X sea estrictamente menor que x, lo cual no incluye al valor x.
$P(X \geq x) = \sum_x P_X(x)$	porque la probabilidad de que $X \geq x$ es la probabilidad de que X valga x o cualquier valor mayor que x.
$P(X > x) = \left(\sum_x P_X(x) \right) - P(X = x)$	es igual que la anterior pero le restamos $P(X = x)$ porque estamos pidiendo que X sea estrictamente mayor que x, lo cual no incluye al valor x.
$P(a \leq X \leq b) = \sum_a^b P_X(x)$	porque la probabilidad de que X esté entre a y b es la probabilidad de que valga a, el valor siguiente de a, el siguiente, ..., ó b.

Si X es continua:

$P(X = x) = 0$	porque la probabilidad de un punto en un continuo es cero.
$P(X \leq x) = \int_{-\infty} f_X(x) dx$	porque la probabilidad de que $X \leq x$ es la probabilidad de que X caiga en el intervalo $(-\infty; x]$
$P(X < x) = P(X \leq x)$	porque $P(X < x) = P(X \leq x) - P(X = x) = P(X \leq x) - 0 = P(X \leq x)$ También podemos verlo porque una integral definida no tiene en cuenta si se incluyen o no los extremos del intervalo de integración. Observemos que esta igualdad entre $P(X \leq x)$ y $P(X < x)$ solo sucede con las variables aleatorias continuas, porque un determinado punto tiene probabilidad cero. En las variables aleatorias discretas, los puntos no tienen probabilidad cero.

$P(X \geq x) = \int_x^{+\infty} f_X(x) dx$	porque la probabilidad de que $X \leq x$ es la probabilidad de que X valga x o cualquier valor mayor que x .
$P(X > x) = P(X \geq x)$	por razones análogas al caso de $P(X < x)$
$P(a \leq X \leq b) = \int_a^b f_X(x) dx$	porque la probabilidad de que X esté entre a y b es la suma de todos los diferenciales de probabilidad en el intervalo $[a;b]$

Función de distribución de probabilidad

(también conocida como función de distribución de probabilidad acumulada a izquierda)

Esta función se nota con la F mayúscula tanto para las distribuciones discretas como continuas. Vale $F_X(x) = P(X \leq x)$. Se la llama función de distribución acumulada porque indica la probabilidad "acumulada" por todos los valores con probabilidad no nula desde $-\infty$ hasta x .

$$F_X(x) = \sum_{x=-\infty} P_X(x)$$

- Para una variable aleatoria discreta:

$$F_X(x) = \int_{-\infty}^x f_X(x) dx$$

- Para una variable aleatoria continua:

Para ser función de distribución de probabilidad, $F_X(x)$ debe cumplir:

$$1) \lim_{x \rightarrow -\infty} F_X(x) = 0$$

porque la probabilidad de que X sea menor que $-\infty$ es cero.

$$2) \lim_{x \rightarrow \infty} F_X(x) = 1$$

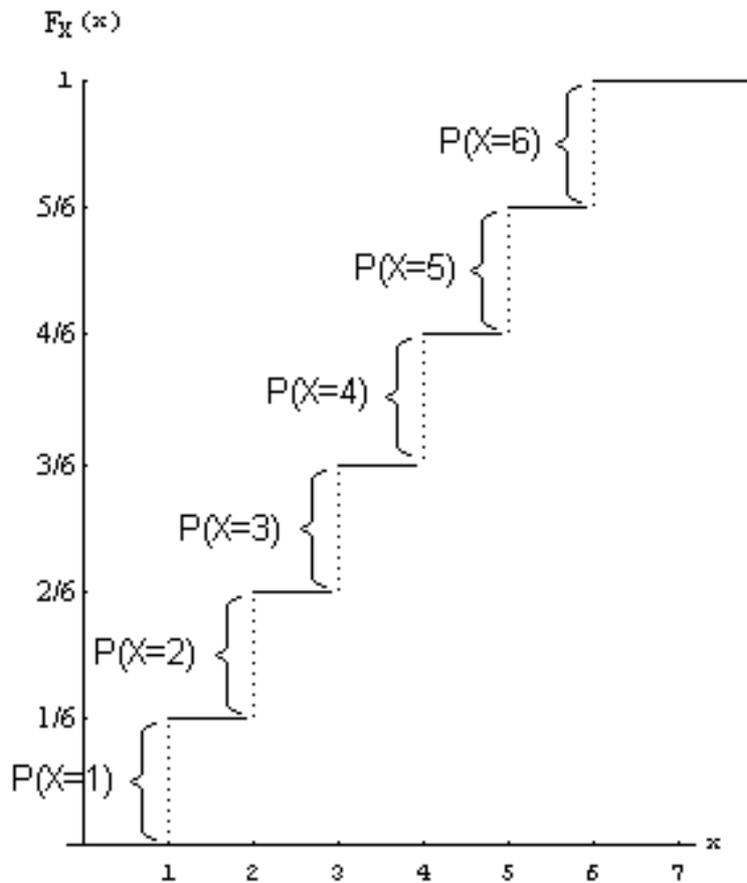
porque la probabilidad de que X sea menor que ∞ es 1.

3) $F_X(x)$ monótonamente creciente (es decir, nunca puede ser decreciente). porque la probabilidad de que X sea menor que un cierto valor no puede ser menor a la probabilidad de que X sea menor a valor mayor que él. Ejemplo: $P(X \leq 5)$ no puede ser nunca mayor que $P(X \leq 6)$. A lo sumo podrá ser igual.

Formalmente: si $b > a$, entonces $F_X(b) \geq F_X(a)$

$$4) \lim_{h \rightarrow 0} F_X(x+h) = F_X(x) \quad (\text{es decir, } F_X(x) \text{ es continua por derecha})$$

esto lo analizaremos por separado para variables aleatorias discretas y continuas:



- Variables aleatorias discretas:

La probabilidad acumulada comienza siendo cero (en $-\infty$). Sigue siendo cero hasta que encuentra el primer valor con probabilidad no nula. A partir de ese valor, la probabilidad acumulada es la probabilidad de ese primer punto. Dicha probabilidad acumulada se mantiene igual, hasta que se llega al segundo punto con probabilidad no nula. A partir de ese punto, la probabilidad acumulada vale la suma de las probabilidades de esos dos puntos. Y así sucesivamente hasta llegar al último valor con probabilidad no nula, a partir del cual la probabilidad acumulada vale uno.

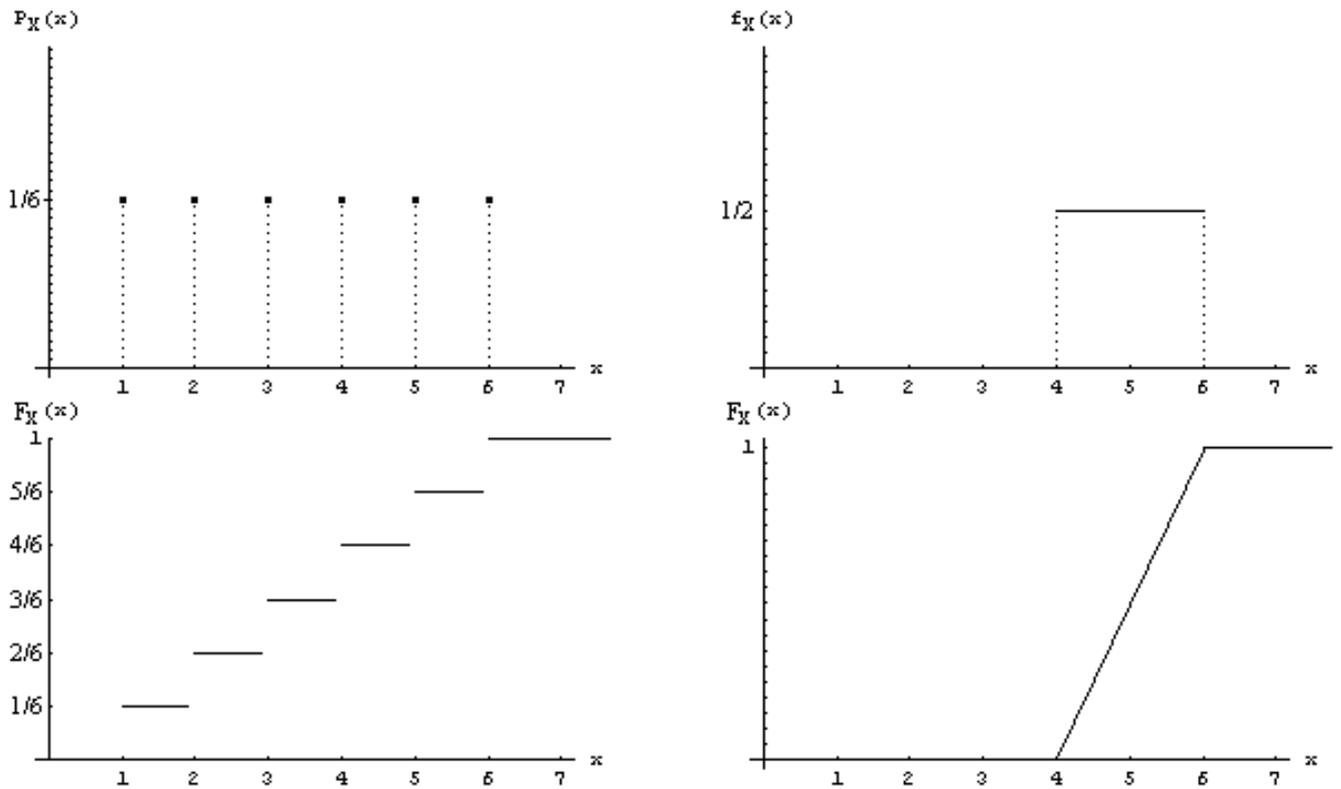
Observemos que el "salto" dado por la función de distribución acumulada en cada punto es igual a la probabilidad de ese punto (porque esa probabilidad es lo que se "agrega" a la suma acumulada a partir de ese punto).

Esto nos muestra que la función tendrá una discontinuidad por izquierda en cada punto con probabilidad no nula, porque la función da el "salto". Pero por derecha es continua, porque a partir de cada valor con probabilidad no nula (incluyendo el valor) la función vale lo mismo (hasta llegar al próximo punto con probabilidad no nula).

- Variables aleatorias continuas:

Como lo que se va sumando en cada punto son diferenciales de probabilidad, entonces la función de distribución acumulada resulta siempre continua, tanto por derecha como por izquierda. Incluso algunos autores definen que una variable aleatoria es continua \Leftrightarrow su función de distribución acumulada es continua.

Observemos ahora un gráfico que nos permite obtener una idea intuitiva de la relación entre la función de densidad y la función de distribución:



Cálculo de probabilidades con la función de distribución

Es justamente para el cálculo de probabilidades que se puede apreciar la utilidad de la función de distribución. Como ésta tiene ya incluida la sumatoria o integral, para calcular probabilidades no hace falta calcular ninguna sumatoria ni integral, justamente porque éstas ya están hechas. Es decir, si por ejemplo queremos saber $P(X \leq 5)$, y conocemos la función de distribución, no necesitaremos hacer ninguna sumatoria ni integral, porque el resultado es directamente $F_X(5)$.

Si X es discreta:

$P(X \leq x) = F_X(x)$	porque $F_X(x)$ es directamente $P(X \leq x)$
$P(X < x) = F_X(x) - P(X = x)$	es igual que la anterior pero le restamos $P(X = x)$ porque estamos pidiendo que X sea estrictamente menor que x, lo cual no incluye al valor x. Si no restáramos $P(X = x)$, estaríamos acumulando probabilidad que no corresponde.
$P(X \geq x) = 1 - P(X < X)$	porque como la probabilidad de que ocurra un suceso o su complemento vale 1, entonces $P(X \leq x) + P(X > x) = 1$
$P(X > x) = 1 - F_X(x)$	justificación análoga a la anterior

$$P(a \leq X \leq b) = F_x(b) - F_x(a) + P(X = a)$$

La diferencia de lo acumulado hasta b y lo acumulado hasta a, es lo que se acumula entre a y b. A eso hay que sumarle $P(X = a)$ porque como $F_x(a)$ incluye la probabilidad en a, al restarla estamos omitiendo en el resultado dicha probabilidad. (*)

(*) Pensemos en el ejemplo del dado:

$$\begin{aligned} P(2 \leq X \leq 5) &= F_x(5) - F_x(2) + P(X = 2) = \\ &= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) - (P(X = 1) + P(X = 2)) + P(X = 2) = \\ &= P(X = 3) + P(X = 4) + P(X = 5) + P(X = 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \end{aligned}$$

y eso es lo que buscábamos.

Además observemos que si a no pertenece al rango de X, la fórmula sigue siendo válida, puesto que $P(X = a) = 0$.

Si X es continua:

$P(X \leq x) = F_x(x)$	porque $F_x(x)$ es directamente $P(X \leq x)$
$P(X < x) = F_x(x)$	porque $P(X < x) = P(X \leq x) - P(X = x) = P(X \leq x) - 0 = P(X \leq x)$ Esto solamente ocurre para variables aleatorias continuas
$P(X > x) = 1 - F_x(x)$	porque como la probabilidad de que ocurra un suceso o su complemento vale 1, entonces $P(X \leq x) + P(X > x) = 1$
$P(X > x) = 1 - F_x(x)$	justificación análoga
$P(a \leq X \leq b) = F_x(b) - F_x(a)$	En este caso no necesitamos sumar $P(X = a)$ porque dicha probabilidad vale 0.

Método para construir la función de distribución a partir de la función de densidad

Como se dijo antes, la función de distribución resulta útil cuando se necesitan calcular muchas probabilidades de una misma variable aleatoria, porque en ese caso sólo se resuelve la sumatoria o integral una vez (al construir la función de distribución) y luego solamente se evalúa dicha función donde corresponda y se obtienen todas las probabilidades buscadas.

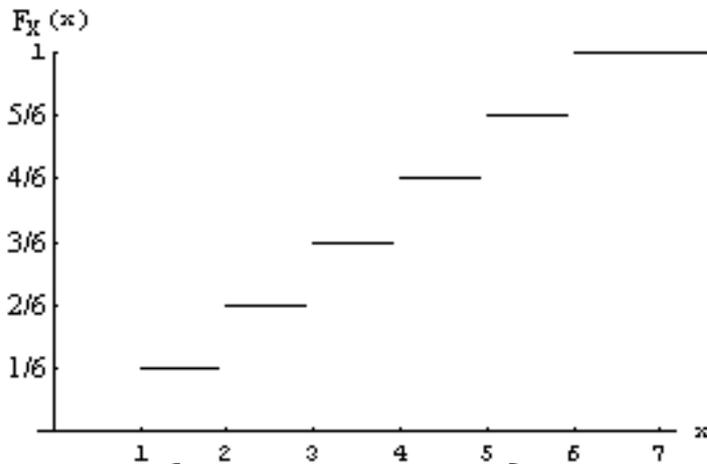
Pero si necesitamos solamente una o dos probabilidades, no vale la pena construir la función de distribución. En tal caso, sólo construiremos la función de distribución si nos la piden explícitamente. Si no nos la piden, no la construiremos, y sumaremos o integraremos la función de densidad para obtener las probabilidades.

Para variables aleatorias discretas:

Cuando la variable es discreta el método es simple. Si los puntos en los que la variable tiene probabilidad no nula son x_1, x_2, \dots, x_n , la función de distribución queda así:

$$F_X(x) = \left\{ \begin{array}{ll} 0 & x < x_1 \\ P_X(x_1) & x_1 \leq x < x_2 \\ P_X(x_1) + P_X(x_2) & x_2 \leq x < x_3 \\ P_X(x_1) + P_X(x_2) + P_X(x_3) & x_3 \leq x < x_4 \\ \dots & \dots \\ 1 & x \geq x_n \end{array} \right.$$

Obviamente en la primera columna pondremos los valores correspondientes a cada probabilidad y haremos la suma. Por ejemplo, en el caso del dado, la función queda así:



$$F_X(x) = \left\{ \begin{array}{ll} 0 & x < 1 \\ 1/6 & 1 \leq x < 2 \\ 2/6 & 2 \leq x < 3 \\ 3/6 & 3 \leq x < 4 \\ 4/6 & 4 \leq x < 5 \\ 5/6 & 5 \leq x < 6 \\ 1 & x \geq 6 \end{array} \right.$$

Para variables aleatorias continuas:

Cuando la variable aleatoria cuya función de distribución queremos hallar es continua, el método es un poco más complejo.

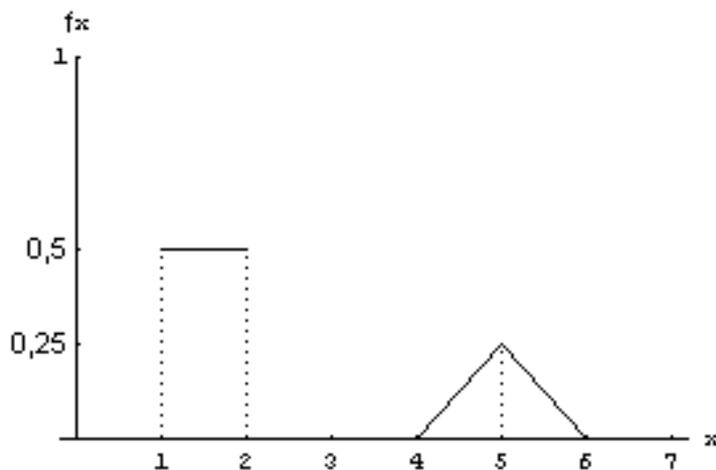
Tomaremos la función de densidad y haremos una lista de los puntos que dividen sus ramas. Luego el dominio nos quedará dividido en intervalos. Trabajaremos intervalo por intervalo, de izquierda a derecha. Usaremos una variable que

llamaremos "a". Inicialmente, $a = 0$.

Para cada intervalo tendremos una definición de f_x , y los extremos del intervalo, que llamaremos x_1 y x_2 . Con cada intervalo haremos lo siguiente:

- 1) Encontrar una primitiva de lo que valga f_x en el intervalo, y la llamarla Φ_x .
- 2) $C = a - \Phi_x(x_1)$
- 3) $F_x(x)$ para el intervalo actual vale: $F_x(x) = \Phi_x(x) + C$
- 4) Actualizar $a = F_x(x_2)$
- 5) Procesar el siguiente intervalo.

Luego de procesados todos los intervalos, la $F_x(x)$ quedará formada como la función que en cada rama valdrá lo que determinamos para cada intervalo, entre los límites x_1 y x_2 de cada intervalo.



Ejemplo:

$$f_x(x) = \left\{ \begin{array}{ll} 1/2 & 1 \leq x \leq 2 \\ \frac{x}{2} - 2 & 4 \leq x \leq 5 \\ 3 - \frac{x}{2} & 5 < x \leq 6 \\ 0 & \forall \text{ otro } x \end{array} \right.$$

Podemos verificar que efectivamente es una función de densidad porque nunca se hace negativa y su integral entre $-\infty$ y $+\infty$ vale 1.

Hallamos los puntos que dividen las ramas: 1, 2, 4, 5, 6 (quedan 6 intervalos).

Tomamos $a = 0$.

Estudiamos todos los intervalos:

- Intervalo $-\infty < x < 1$

$$f_x = 0$$

$$\Phi_x = 0$$

$$C = a - \Phi_x(x_1) = a - \Phi_x(-\infty) = 0 - 0 = 0$$

$$F_x(x) = \Phi_x(x) + C = 0$$

$$a = F_x(x_2) = 0$$

• Intervalo $1 < x < 2$

$$f_x = 0,5$$

$$\Phi_x = 0,5x$$

$$C = a - \Phi_x(x_1) = a - \Phi_x(1) = 0 - 0,5 = -0,5$$

$$F_x(x) = \Phi_x(x) + C = 0,5x - 0,5$$

$$a = F_x(x_2) = F_x(2) = 0,5$$

• Intervalo $2 < x < 4$

$$f_x = 0$$

$$\Phi_x = 0$$

$$C = a - \Phi_x(x_1) = a - \Phi_x(2) = 0,5 - 0$$

$$F_x(x) = \Phi_x(x) + C = 0 + 0,5 = 0,5$$

$$a = F_x(x_2) = F_x(4) = 0,5$$

• Intervalo $4 < x < 5$

$$f_x = x/2 - 2$$

$$\Phi_x = x^2/4 - 2x$$

$$C = a - \Phi_x(x_1) = a - \Phi_x(4) = 0,5 - (-4) = 4,5$$

$$F_x(x) = \Phi_x(x) + C = x$$



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/processtext/>

$$x^2/4 - 2x + 4,5$$

$$a = F_x(x_2) = F_x(5) = 0,75$$

• Intervalo $5 < x < 6$

$$f_x = 3 - x/2$$

$$\Phi_x = 3x - x^2/4$$

$$C = a - \Phi_x(x_1) = a - \Phi_x(5) = 0,75 - 8,75 = -8$$

$$F_x(x) = \Phi_x(x) + C = 3x - x^2/4 - 8$$

$$a = F_x(x_2) = F_x(6) = 1$$

• Intervalo $6 < x < +\infty$

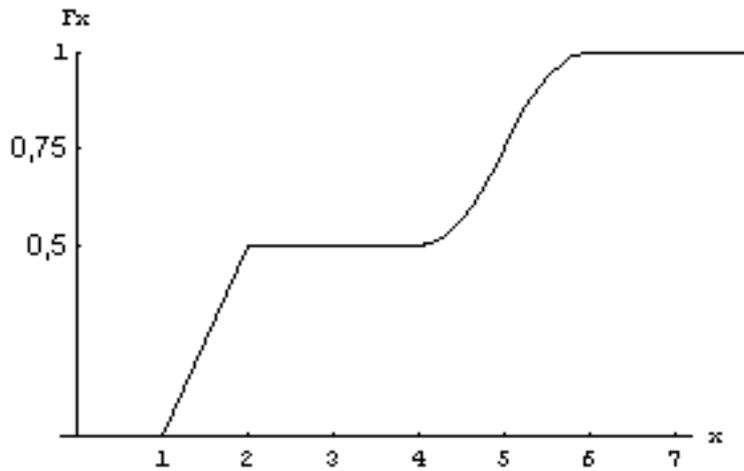
$$f_x = 0$$

$$\Phi_x = 0$$

$$C = a - \Phi_x(x_1) = a - \Phi_x(6) = 1 - 0 = 1$$

$$F_x(x) = \Phi_x(x) + C = 0 + 1 = 1$$

$$a = F_x(x_2) = F_x(+\infty) = 1$$



Y luego juntamos todas las ramas para armar la $F_x(x)$:

$$F_x(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{2}(x-1) & 1 \leq x < 2 \\ \frac{1}{2} & 2 \leq x < 4 \\ \frac{1}{4}x^2 - 2x + 4,5 & 4 \leq x < 5 \\ -\frac{1}{4}x^2 + 3x - 8 & 5 \leq x < 6 \\ 1 & x \geq 6 \end{cases}$$

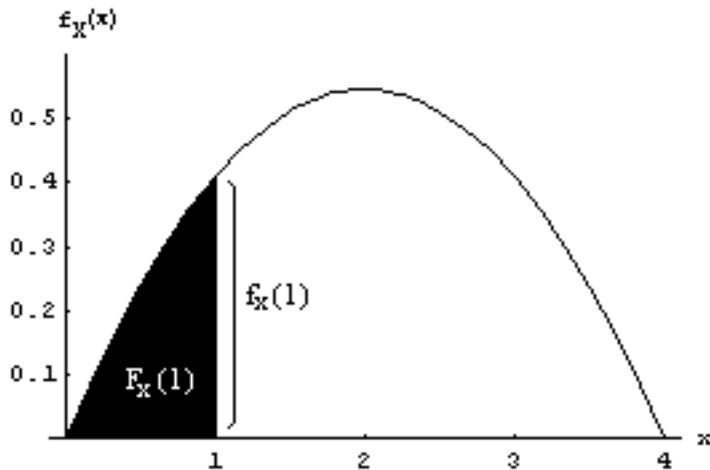
Vemos que la función de distribución de una variable continua es continua.

Método para obtener la función de densidad a partir de la función de distribución

Para variables aleatoria discretas:

Reconoceremos que una función de distribución pertenece a una variable aleatoria discreta porque es constante en todo \mathcal{R} salvo en una cantidad finita o infinita numerable de puntos, en los cuales tiene "saltos". La técnica para obtener la función de densidad, dada la función de distribución, es bastante simple, y consiste en definir la función $P_x(x)$ con valor nulo en todo \mathcal{R} salvo en los puntos en los que la función de distribución tiene "saltos", en los cuales $P_x(x)$ vale la altura del salto. Por ejemplo, en el caso del dado, tenemos:

$$F_X(x) = \begin{cases} 0 & x < 1 \\ 1/6 & 1 \leq x < 2 \\ 2/6 & 2 \leq x < 3 \\ 3/6 & 3 \leq x < 4 \\ 4/6 & 4 \leq x < 5 \\ 5/6 & 5 \leq x < 6 \\ 1 & x \geq 6 \end{cases} \Rightarrow P_X(x) = \begin{cases} 1/6 - 0 & x = 1 \\ 2/6 - 1/6 & x = 2 \\ 3/6 - 2/6 & x = 3 \\ 4/6 - 3/6 & x = 4 \\ 5/6 - 4/6 & x = 5 \\ 6/6 - 5/6 & x = 6 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} 1/6 & x = 1 \\ 1/6 & x = 2 \\ 1/6 & x = 3 \\ 1/6 & x = 4 \\ 1/6 & x = 5 \\ 1/6 & x = 6 \\ 0 & \forall \text{ otro } x \end{cases}$$



Para variables aleatorias continuas:

$$F_X(x) = \int_{-\infty}^x f_X(x) dx$$

Como $\int_{-\infty}^x f_X(x) dx$, también se cumple que $f_X(x)$ es la derivada de $F_X(x)$ respecto de x , porque $F_X(x)$ es el área bajo la curva de $f_X(x)$, y al mismo tiempo $f_X(x)dx$ son los diferenciales de probabilidad que se acumulan al integrar para encontrar la $F_X(x)$. Esto lo podremos ver más claramente en el gráfico.

Cada rama de $f_X(x)$ se puede obtener derivando la rama correspondiente de $F_X(x)$, en un procedimiento más sencillo que obtener $F_X(x)$ a partir de $f_X(x)$ ya que no es necesario buscar constantes para que la función resulte continua.

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{2}(x-1) & 1 \leq x < 2 \\ \frac{1}{2} & 2 \leq x < 4 \\ \frac{1}{4}x^2 - 2x + 4,5 & 4 \leq x < 5 \\ -\frac{1}{4}x^2 + 3x - 8 & 5 \leq x < 6 \\ 1 & x \geq 6 \end{cases} \Rightarrow f_X(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{2} & 1 \leq x < 2 \\ 0 & 2 \leq x < 4 \\ \frac{1}{2}x - 2 & 4 \leq x < 5 \\ -\frac{1}{2}x + 3 & 5 \leq x < 6 \\ 0 & x \geq 6 \end{cases} = \begin{cases} \frac{1}{2} & 1 \leq x \leq 2 \\ \frac{x}{2} - 2 & 4 \leq x \leq 5 \\ 3 - \frac{x}{2} & 5 < x \leq 6 \\ 0 & \forall \text{ otro } x \end{cases}$$

Función de distribución acumulada a derecha

Esta función se nota con la G mayúscula tanto para las distribuciones discretas

como continuas. Vale $G_X(x) = P(X \geq x)$. Se la llama función de distribución acumulada a derecha porque indica la probabilidad "acumulada" por todos los valores con probabilidad no nula desde x hasta $+\infty$.

Como el lector podrá imaginar, es completamente análoga a la función de distribución acumulada a izquierda $F_X(x)$, y cumple con las siguientes propiedades:

1) $\lim_{x \rightarrow -\infty} G_X(x) = 1$

porque la probabilidad de que X sea mayor que $-\infty$ es uno.

2) $\lim_{x \rightarrow \infty} G_X(x) = 0$

porque la probabilidad de que X sea mayor que ∞ es 0.

3) $G_X(x)$ monótonamente decreciente (es decir, nunca puede ser creciente). porque la probabilidad de que X sea mayor que un cierto valor no puede ser mayor a la probabilidad de que X sea mayor a valor menor que él. Ejemplo: $P(X \geq 5)$ no puede ser nunca menor que $P(X \geq 6)$. A lo sumo podrá ser igual.

Formalmente: si $b < a$, entonces $F_X(b) \leq F_X(a)$

4) $\lim_{h \rightarrow 0} G_X(x-h) = G_X(x)$ (es decir, $G_X(x)$ es continua por izquierda)

Análogamente a como ocurría con la $F_X(x)$, si X es continua entonces $G_X(x)$ es continua, y si X es discreta entonces $G_X(x)$ es continua por izquierda y discontinua por derecha.

Observación:

Si X es discreta, entonces $F_X(x) + G_X(x) = 1 + P_X(x)$ porque al sumar la $F_X(x)$ y la $G_X(x)$ estamos contando 2 veces $P(X = x)$.

Si X es continua, entonces $F_X(x) + G_X(x) = 1$ porque $P(X = x) = 0$.

Problemas típicos

1) Determine si las siguientes funciones son de densidad de probabilidad:

a)
$$P_X(x) = \begin{cases} 0,5 & x = 2 \\ 0,3 & x = 3 \\ 0,3 & x = 6 \\ 0 & \forall \text{ otro } x \end{cases}$$

b)
$$P_X(x) = \begin{cases} 0,5 & x = 2 \\ -0,2 & x = 3 \\ 0,7 & x = 6 \\ 0 & \forall \text{ otro } x \end{cases}$$

c)
$$P_X(x) = \begin{cases} 0,5 & x = 2 \\ 0,3 & x = 3 \\ 0,2 & x = 6 \\ 0 & \forall \text{ otro } x \end{cases}$$

d)
$$P_X(x) = \begin{cases} 0,5 & x = 2 \\ -0,3 & x = 3 \\ 0,2 & x = 6 \\ 0 & \forall \text{ otro } x \end{cases}$$

e)
$$P_X(x) = \begin{cases} 0,4 & x = -1 \\ 0,1 & x = 1 \\ 0,3 & x = 2 \\ 0,2 & x = 3 \\ 0 & \forall \text{ otro } x \end{cases}$$

Resolución:

- a) No. No suma 1.
- b) No. Suma 1 pero asigna probabilidades negativas.
- c) Si. Suma 1 y es $\geq 0 \forall x$
- d) No. No suma 1 y asigna probabilidades negativas.
- e) Si. Suma 1 y es $\geq 0 \forall x$

2) La variable aleatoria discreta X está distribuida según:

$$P_X(x) = \begin{cases} 0,1 & x = 1 \\ 0,15 & x = 2 \\ 0,1 & x = 4 \\ 0,2 & x = 5 \\ 0,45 & x = 7 \\ 0 & \forall \text{ otro } x \end{cases}$$

Indique la probabilidad de que X:

- a) sea 4.
- b) sea menor o igual a 4.
- c) sea a lo sumo 4.
- d) sea menor a 4.
- e) sea mayor o igual a 4.
- f) sea como mínimo 4.
- g) sea mayor a 4.
- h) esté entre 3 y 6.
- i) sea menor que 4, sabiendo que es menor que 5.

Resolución:

- a) $P(X = 4) = 0,1$
 b) $P(X \leq 4) = P(X = 1) + P(X = 2) + P(X = 4) = 0,35$
 c) "a lo sumo 4" y "como máximo 4" significan lo mismo que "menor o igual a 4".
 Están pidiendo lo mismo que en el punto b.
 d) $P(X < 4) = P(X = 1) + P(X = 2) = 0,25$
 e) $P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 7) = 0,75$
 f) "como mínimo 4" y "al menos 4" significan lo mismo que "mayor o igual a 4".
 Están pidiendo lo mismo que en el punto e.
 g) $P(X > 4) = P(X = 5) + P(X = 7) = 0,65$
 h) $P(3 \leq X \leq 6) = P(X = 4) + P(X = 5) = 0,3$

$$P\left(\frac{X < 4}{X < 5}\right) = \frac{P(X < 4 \wedge X < 5)}{P(X < 5)}$$

i) Nos piden

Vemos que la condición del numerador se puede reducir a $P(X < 4)$, porque el suceso $\{X < 4\}$ es un subconjunto de $\{X < 5\}$, con lo cual $\{X < 4\} \Rightarrow \{X < 5\}$

Luego:

$$P\left(\frac{X < 4}{X < 5}\right) = \frac{P(X < 4 \wedge X < 5)}{P(X < 5)} = \frac{P(X < 4)}{P(X < 5)} = \frac{P(X = 1) + P(X = 2)}{P(X = 1) + P(X = 2) + P(X = 4)} = \frac{0,25}{0,35} = 0,71$$

3) Halle las funciones de densidad y distribución de probabilidad de la cantidad de caras obtenidas al tirar 2 monedas.

Resolución:

Como vimos en los ejemplos del capítulo 1, al tirar 2 monedas la probabilidad de no obtener ninguna cara es $1/4$, la probabilidad de obtener 2 caras también es $1/4$, y la probabilidad de obtener una cara es $1/2$.

Entonces si X es la cantidad de caras obtenidas al tirar 2 monedas, la función de densidad queda:

$$P_X(x) = \begin{cases} 0,25 & x = 0 \\ 0,5 & x = 1 \\ 0,25 & x = 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

Y la función de distribución queda:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ P(X = 0) & 0 \leq x < 1 \\ P(X = 0) + P(X = 1) & 1 \leq x < 2 \\ P(X = 0) + P(X = 1) + P(X = 2) & x \geq 2 \end{cases} = \begin{cases} 0 & x < 0 \\ 0,25 & 0 \leq x < 1 \\ 0,25 + 0,5 & 1 \leq x < 2 \\ 0,25 + 0,5 + 0,25 & x \geq 2 \end{cases} = \begin{cases} 0 & x < 0 \\ 0,25 & 0 \leq x < 1 \\ 0,75 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

4) La función de distribución de la variable aleatoria discreta X es:

$$F_X(x) = \begin{cases} 0 & x < -1 \\ 0,4 & -1 \leq x < 1 \\ 0,5 & 1 \leq x < 2 \\ 0,8 & 2 \leq x < 3 \\ 1 & x > 3 \end{cases}$$

Halle la función de densidad de probabilidad de X.

Resolución:

Como se explicó antes, para obtener la función de densidad de una variable aleatoria discreta, dada la función de distribución, se define la función $P_X(x)$ con valor nulo en todo \mathfrak{R} salvo en los puntos en los que la función de distribución tiene "saltos", en los cuales $P_X(x)$ vale la altura del salto.

$$F_X(x) = \begin{cases} 0 & x < -1 \\ 0,4 & -1 \leq x < 1 \\ 0,5 & 1 \leq x < 2 \\ 0,8 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases} \Rightarrow P_X(x) = \begin{cases} 0,4 - 0 & x = -1 \\ 0,5 - 0,4 & x = 1 \\ 0,8 - 0,5 & x = 2 \\ 1 - 0,8 & x = 3 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} 0,4 & x = -1 \\ 0,1 & x = 1 \\ 0,3 & x = 2 \\ 0,2 & x = 3 \\ 0 & \forall \text{ otro } x \end{cases}$$

5) Determine si las siguientes funciones son de densidad de probabilidad:

a) $f_X(x) = \begin{cases} 0,5 & 0 \leq x \leq 3 \\ 0 & \forall \text{ otro } x \end{cases}$

b) $f_X(x) = \begin{cases} \frac{3}{4}(x^2 - 4x + 3) & 0 \leq x \leq 4 \\ 0 & \forall \text{ otro } x \end{cases}$

c) $f_X(x) = \begin{cases} \frac{3}{16}x^2 & -2 \leq x \leq 2 \\ 0 & \forall \text{ otro } x \end{cases}$

Resolución:

a) $\int_{-\infty}^{+\infty} f_X(x) dx = \int_0^3 0,5 dx = 1,5 \neq 1 \Rightarrow$ no es función de densidad

b) $\int_{-\infty}^{+\infty} f_X(x) dx = \int_0^4 \frac{3}{4}(x^2 - 4x + 3) dx = 1$, pero no se cumple $f_X(x) \geq 0 \forall x$, puesto que para $1 < x < 3$, $f_X(x) < 0 \Rightarrow$ no es función de densidad

c) $\int_{-\infty}^{+\infty} f_X(x) dx = \int_0^2 \frac{3}{16}x^2 dx = 1$ y se cumple $f_X(x) \geq 0 \forall x \Rightarrow$ es función de densidad.

6) La variable aleatoria continua X está distribuida según:

$$f_X(x) = \begin{cases} \frac{1}{3} & 1 \leq x \leq 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

Indique la probabilidad de que X:

- a) sea 3.
- b) sea menor o igual a 3.
- c) sea a lo sumo 3.
- d) sea menor a 3.
- e) sea mayor o igual a 3.
- f) sea como mínimo 3.
- g) sea mayor a 4.
- h) esté entre 3 y 6.
- i) sea menor que 2, sabiendo que es menor que 3.
- j) sea menor que 3.5, sabiendo que es mayor que 1.5.
- k) Sean los sucesos A y B:
 A: $X < 2$
 B: $X > 3$

Determine si A y B son independientes.

Resolución:

a) $P(X = 3) = 0$ (es un punto en un continuo).

$$P(X \leq 3) = \int_{-\infty}^3 f_X(x) dx = \int_1^3 \frac{1}{3} dx = \frac{2}{3}$$

b)

c) "a lo sumo 3" y "como máximo 3" significan lo mismo que "menor o igual a 3". Están pidiendo lo mismo que en el punto b.

d) $P(X < 3) = P(X \leq 3) = \frac{2}{3}$ como en el punto b.

$$P(X \geq 3) = \int_3^{+\infty} f_X(x) dx = \int_3^4 \frac{1}{3} dx = \frac{1}{3}$$

e)

También se podría haber hecho $P(X \geq 3) = 1 - P(X < 3) = \frac{1}{3}$.

f) "como mínimo 3" y "al menos 3" significan lo mismo que "mayor o igual a 3". Están pidiendo lo mismo que en el punto e.

g) $P(X > 3) = P(X \geq 3) = \frac{1}{3}$ como en el punto e.

$$P(3 \leq X \leq 6) = \int_3^6 f_X(x) dx = \int_3^4 f_X(x) dx + \int_4^6 f_X(x) dx = \int_3^4 \frac{1}{3} dx + \int_4^6 0 dx = \frac{1}{3}$$

h)

$$P\left(\frac{X < 2}{X < 3}\right) = \frac{P(X < 2 \wedge X < 3)}{P(X < 3)}$$

i) Nos piden

Vemos que la condición del numerador se puede reducir a $P(X < 2)$, porque el suceso $\{X < 2\}$ es un subconjunto de $\{X < 3\}$, con lo cual $\{X < 2\} \Rightarrow \{X < 3\}$

Luego:

$$P\left(\frac{X < 2}{X < 3}\right) = \frac{P(X < 2 \wedge x < 3)}{P(X < 3)} = \frac{P(X < 2)}{P(X < 3)} = \frac{\int_1^2 \frac{1}{3} dx}{\int_1^3 \frac{1}{3} dx} = \frac{1}{2}$$

$$P\left(\frac{X < 3.5}{X > 1.5}\right) = \frac{P(X < 3.5 \wedge X > 1.5)}{P(X > 1.5)} = \frac{\int_{1.5}^{3.5} \frac{1}{3} dx}{\int_{1.5}^{\infty} \frac{1}{3} dx} = 0.8$$

j) Nos piden

k) A, B son independientes $\Leftrightarrow P(A) \cdot P(B) = P(A \cap B)$

$$P(A) = P(X < 2) = \int_1^2 \frac{1}{3} dx = \frac{1}{3}$$

$$P(B) = P(X > 3) = \int_3^{\infty} \frac{1}{3} dx = \frac{1}{3}$$

$$P(A) \cdot P(B) = \frac{1}{9}$$

$$P(A \cap B) = P(X < 2 \wedge x > 3) = 1 - P(2 < X < 3) = 1 - \int_2^3 \frac{1}{3} dx = 1 - \frac{1}{3} = \frac{2}{3} \neq \frac{1}{9} \Rightarrow$$

\Rightarrow no son independientes.

Este resultado lo podemos entender intuitivamente si pensamos que si $X < 2$, entonces se ve afectada la probabilidad de que $X > 3$ (se hace cero) y viceversa.

7) Halle la función de distribución de la variable aleatoria X distribuida según:

$$f_X(x) = \begin{cases} x^2 & 0 \leq x \leq 1 \\ \frac{1}{3} & 3 \leq x \leq 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

Resolución:

Según el método explicado, comenzamos por listar los puntos que dividen las ramas de $f_X(x)$: 0, 1, 3, 5. Son 4 puntos, con lo cual \mathfrak{R} nos queda partido en 5 intervalos.

Tomamos $a = 0$.

• Intervalo $-\infty < x < 0$

$$f_X = 0$$

$$\Phi_X = 0$$

$$C = a - \Phi_X(x_1) = 0 - 0 = 0$$

$$F_X(x) = \Phi_X(x) + C = 0 + 0 = 0$$

$$a = F_X(x_2) = F_X(0) = 0$$

• Intervalo $0 < x < 1$

$$f_X = x^2$$

$$\Phi_x = x^3/3$$

$$C = a - \Phi_x(x_1) = a - \Phi_x(0) = 0 - 0 = 0$$

$$F_x(x) = \Phi_x(x) + C = x^3/3 + 0 = x^3/3$$

$$a = F_x(x_2) = F_x(1) = 1/3$$

• Intervalo $1 < x < 3$

$$f_x = 0$$

$$\Phi_x = 0$$

$$C = a - \Phi_x(x_1) = a - \Phi$$



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/processtext/>

$$x(1) = 1/3 - 0 = 1/3$$

$$F_x(x) = \Phi_x(x) + C = 0 + 1/3 = 1/3$$

$$a = F_x(x_2) = F_x(3) = 1/3$$

• Intervalo $3 < x < 5$

$$f_x = 1/3$$

$$\Phi_x = x/3$$

$$C = a - \Phi_x(x_1) = a - \Phi_x(3) = 1/3 - 1 = -2/3$$

$$F_x(x) = \Phi_x(x) + C = x/3 - 2/3$$

$$a = F_x(x_2) = F_x(5) = 1$$

• Intervalo $5 < x < +\infty$

$$f_x = 0$$

$$\Phi_x = 0$$

$$C = a - \Phi_x(x_1) = a - \Phi_x(5) = 1 - 0 = 1$$

$$F_x(x) = \Phi_x(x) + C = 0 + 1 = 1$$

$$a = F_x(x_2) = F_x(+\infty) = 1$$

Y luego juntamos todas las ramas para armar la $F_x(x)$:

$$F_x(x) = \left\{ \begin{array}{ll} 0 & x < 0 \\ \frac{1}{3}x^3 & 0 \leq x \leq 1 \\ \frac{1}{3} & 1 \leq x \leq 3 \\ \frac{x-2}{3} & 3 \leq x \leq 5 \\ 1 & x \geq 5 \end{array} \right\}$$

8) Dada la siguiente función de distribución de probabilidad, indique la función de densidad de la distribución:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{8}x^2 & 0 \leq x < 2 \\ \frac{1}{4}x & 2 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

Resolución:

Según el método explicado, se obtiene $f_X(x)$ derivando cada rama de $F_X(x)$:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{8}x^2 & 0 \leq x < 2 \\ \frac{1}{4}x & 2 \leq x < 4 \\ 1 & x \geq 4 \end{cases} = f_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4}x & 0 \leq x < 2 \\ \frac{1}{4} & 2 \leq x < 4 \\ 0 & x \geq 4 \end{cases} = \begin{cases} \frac{1}{4}x & 0 \leq x < 2 \\ \frac{1}{4} & 2 \leq x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

9) Determinar para qué valor de k las siguientes funciones son funciones de densidad de probabilidad:

a)
$$f_X(x) = \begin{cases} kx^2 & 0 \leq x \leq 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

b)
$$f_X(x) = \begin{cases} x^2 & 0 \leq x \leq k \\ 0 & \forall \text{ otro } x \end{cases}$$

Resolución:

a) Una condición necesaria para que una función sea función de densidad es que su integral sobre \mathfrak{R} de 1. Usando esa condición, queda:

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1 \Rightarrow \int_0^2 kx^2 dx = 1 \Rightarrow k \int_0^2 x^2 dx = 1 \Rightarrow k \frac{8}{3} = 1 \Rightarrow k = \frac{3}{8}$$

Vemos que ese valor de k también hace que la función cumpla con la otra condición necesaria, es decir, $f_X(x) \geq 0 \forall x$. Cumplidas esas 2 condiciones, determinamos que con el valor de k hallado, la función es función de densidad de probabilidad.

b) De manera análoga al ejercicio anterior, planteamos:

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1 \Rightarrow \int_0^k x^2 dx = 1 \Rightarrow \frac{1}{3}k^3 = 1 \Rightarrow k^3 = 3 \Rightarrow k = \sqrt[3]{3} = 1,442$$

En este caso también vemos que ese valor de k también hace que la función cumpla con la otra condición necesaria, es decir, $f_X(x) \geq 0 \forall x$. Cumplidas esas 2 condiciones, determinamos que con el valor de k hallado, la función es función de

densidad de probabilidad.



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/process>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 4 de mayo de 2004

Cambio de Variables Aleatorias ó Funciones de Variables Aleatorias

En lo que hace a la probabilidad, el cambio de variables consiste en tomar una variable, que tiene una determinada distribución, y calcular la distribución de una función de esa variable.

Por ejemplo, tenemos una variable aleatoria X , distribuida según f_x , y queremos saber cómo estará distribuida la variable Y , dada por $y(x) = 2x^2 + \text{sen}(x)$. Es decir, a partir de $f_x(x)$ y de $y(x)$, obtenemos $f_y(y)$.

Cambio de Variables Discretas

Se tiene la variable aleatoria discreta X , distribuida según P_x , y se desea obtener la distribución P_y de otra variable aleatoria Y que es una función $\varphi(x)$.

Método para obtener P_y a partir de P_x y $\varphi(x)$

- 1) Inicialmente todos los valores de P_y son cero.
- 2) Se recorren los valores posibles de X , evaluando para cada uno $\varphi(x)$, y sumando la probabilidad al valor de P_y correspondiente.

Justificación

Podemos justificar este proceder diciendo que podemos escribir $P(Y = y)$ como la

suma de todos los $P(X = x_i)$ tales que $\varphi(x_i) = y$. Es decir, en el caso general, para cada valor "y" habrá varios valores "x" que hagan que $\varphi(x) = y$.

$$P(Y = y) = P(X = x_1 \cup X = x_2 \cup \dots \cup X = x_n)$$

Luego, como cada uno de los sucesos que están en la unión son disjuntos (porque si $x_i \neq x_j$ entonces X no puede valer x_i y x_j al mismo tiempo) entonces las uniones se pueden transformar en sumas, y eso es lo que hace el método.

Problemas típicos

1) En un determinado juego de tablero, se avanza el doble de casillas de lo que indique un dado honesto. ¿Cómo se distribuyen las probabilidades de las casillas que se avanzan en un turno cualquiera?

Resolución

Como el dado es honesto, tenemos:

$$P_x = \left\{ \begin{array}{ll} 1/6 & x = 1 \\ 1/6 & x = 2 \\ 1/6 & x = 3 \\ 1/6 & x = 4 \\ 1/6 & x = 5 \\ 1/6 & x = 6 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

$$Y = 2X$$

Como tenemos P_x y $\varphi(x)$, podemos obtener la distribución de Y.

Recorremos los valores posibles de X:

$$x = 1 \rightarrow \varphi(1) = 2 \rightarrow \text{sumamos } 1/6 \text{ a } P(Y = 2)$$

$$x = 2 \rightarrow \varphi(2) = 4 \rightarrow \text{sumamos } 1/6 \text{ a } P(Y = 4)$$

$$x = 3 \rightarrow \varphi(3) = 6 \rightarrow \text{sumamos } 1/6 \text{ a } P(Y = 6)$$

$$x = 4 \rightarrow \varphi(4) = 8 \rightarrow \text{sumamos } 1/6 \text{ a } P(Y = 8)$$

$$x = 5 \rightarrow \varphi(5) = 10 \rightarrow \text{sumamos } 1/6 \text{ a } P(Y = 10)$$

$$x = 6 \rightarrow \varphi(6) = 12 \rightarrow \text{sumamos } 1/6 \text{ a } P(Y = 12)$$

Es decir, hicimos:

$$P(Y = 2) = P(X = 1) = 1/6$$

$$P(Y = 4) = P(X = 2) = 1/6$$

$$P(Y = 6) = P(X = 3) = 1/6$$

$$P(Y = 8) = P(X = 4) = 1/6$$

$$P(Y = 10) = P(X = 5) = 1/6$$

$$P(Y = 12) = P(X = 6) = 1/6$$

Luego:

$$P_Y = \left\{ \begin{array}{ll} 1/6 & y = 2 \\ 1/6 & y = 4 \\ 1/6 & y = 6 \\ 1/6 & y = 8 \\ 1/6 & y = 10 \\ 1/6 & y = 12 \\ 0 & \forall \text{ otro } y \end{array} \right\}$$

2) Una cierta variable aleatoria discreta **X** está distribuida según:

$$P_X = \left\{ \begin{array}{ll} 0.05 & x = -2 \\ 0.03 & x = -1 \\ 0.06 & x = 0 \\ 0.12 & x = 1 \\ 0.26 & x = 2 \\ 0.48 & x = 3 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

Si **Y = |X|**, encuentre la distribución de **Y**

Resolución

Recorremos los valores posibles de **X**:

$x = -2 \rightarrow \Phi(-2) = 2 \rightarrow$ sumamos 0,05 a $P(Y = 2)$

$x = -1 \rightarrow \Phi(-1) = 1 \rightarrow$ sumamos 0,03 a $P(Y = 1)$

$x = 0 \rightarrow \Phi(0) = 0 \rightarrow$ sumamos 0,06 a $P(Y = 0)$

$x = 1 \rightarrow \Phi(1) = 1 \rightarrow$ sumamos 0,12 a $P(Y = 1)$

$x = 2 \rightarrow \Phi(2) = 2 \rightarrow$ sumamos 0,26 a $P(Y = 2)$

$x = 3 \rightarrow \Phi(3) = 3 \rightarrow$ sumamos 0,48 a $P(Y = 3)$

Es decir, hicimos:

$$P(Y = 2) = P(X = -2) + P(X = 2) = 0.31$$

$$P(Y = 1) = P(X = -1) + P(X = 1) = 0.15$$

$$P(Y = 0) = P(X = 0) = 0.06$$

$$P(Y = 3) = P(X = 3) = 0.48$$

Luego:

$$P_Y = \left\{ \begin{array}{ll} 0.06 & y = 0 \\ 0.15 & y = 1 \\ 0.31 & y = 2 \\ 0.48 & y = 3 \\ 0 & \forall \text{ otro } y \end{array} \right\}$$

Cambio de Variables Continuas

Dadas $f_x(x)$ e $y(x)$, podemos obtener $f_Y(y)$ mediante:

$$f_Y(y) = \frac{f_X(x)}{\left| \frac{dy}{dx} \right|}$$

Método para obtener $f_Y(y)$

- 1.a) Asegurarse de que se conocen $f_x(x)$ e $y(x)$.
- 1.b) Construir la función $y'(x)$ para todas las ramas de $y(x)$
- 1.c) Hacer la lista de:
 - los puntos que dividan ramas de $f_x(x)$
 - los puntos que dividan ramas de $y(x)$
 - los puntos en los que $y'(x)$ cambie de signo

Una vez determinados los puntos, el dominio de X nos quedará particionado en un conjunto de intervalos. Ahora le aplicaremos la etapa 2 a CADA UNO de esos intervalos.

2.a) Llamaremos al intervalo actual $x_1 < x < x_2$.

2.b) Determinar si, en el intervalo actual, $\frac{dy}{dx}$ es mayor, menor o igual que cero.

- Si es mayor que cero, tomar $\left| \frac{dy}{dx} \right| = \frac{dy}{dx}$
- Si es menor que cero, tomar $\left| \frac{dy}{dx} \right| = -\frac{dy}{dx}$
- Si es igual a cero, saltar a 2.i

2.c) Dados x_1 y x_2 los extremos del intervalo, obtener y_1 e y_2 mediante $y_1=y(x_1)$ e $y_2=y(x_2)$.

2.d) Determinar la expresión válida de $f_x(x)$ para el intervalo actual.

$$\frac{f_X(x)}{\left| \frac{dy}{dx} \right|}$$

2.e) Escribir, para el intervalo actual, y hacer todas las simplificaciones posibles.

2.f) Determinar la expresión válida de $x(y)$, es decir, y^{-1} , para el intervalo actual.

2.g) Reemplazar en la expresión obtenida en 2.e todas las x que aparezcan por la expresión de x en términos de y obtenida en 2.f.

2.h) La expresión obtenida en 2.g constituye el aporte del intervalo actual a la distribución de la y . Este aporte será válido en el intervalo $y_1 < y < y_2$. Ahora analizaremos el siguiente intervalo de X . Si quedan más intervalos de X por analizar, ir al paso 2.a con el siguiente intervalo. De lo contrario, ir al paso 3.

2.i) Como la derivada en este intervalo da cero, Y no depende de X . Entonces en este intervalo de X , Y es constante. Llamaremos y a ese valor constante, e integraremos $f_X(x)$ para averiguar la probabilidad de que Y asuma ese valor:

$$P(Y = y) = \int_{x_1}^{x_2} f_X(x).dx$$

Luego de concluido el método, nos va a quedar una distribución mixta (explicada en la sección 7 del capítulo 2). Es decir, nos quedará por un lado una f_Y que no cerrará a 1, acompañada de una P_Y . La suma de estas dos funciones nos dará efectivamente una distribución que cerrará a 1.

Lo que acabamos de calcular es una rama de P_Y . Ahora pasamos al siguiente intervalo de X . Ir a 2.a

3) Luego de analizados todos los intervalos de X , nos queda una lista de intervalos de Y (que pueden estar superpuestos), junto con expresión en función de Y , para cada uno de esos intervalos de Y . A esas expresiones en función de Y las llamaremos aportes.

3.a) Hacer la lista de los y_1 e y_2 de cada uno de los intervalos de Y que obtuvimos.

3.b) Ahora vamos a la $f_Y(y)$. Los puntos obtenidos en 3.a son los que van a separar las ramas de $f_Y(y)$. Para cada rama, la definición de $f_Y(y)$ será la suma de los aportes cuyos intervalos contengan al intervalo de la rama.

3.c) Hemos obtenido $f_Y(y)$. Fin del método.

Explicación y justificación del método

La fórmula para el cambio de variables involucra a 3 funciones:

- $f_X(x)$
- $y(x)$
- $\left| \frac{dy}{dx} \right|$

Como en el caso general cada una de esas 3 funciones puede tener más de una rama, vamos a particionar el dominio de la X en intervalos para los cuales no cambie la definición de ninguna de las 3 funciones involucradas. Por eso lo dividimos teniendo en cuenta los puntos que dividan las ramas de $f_X(x)$, los que dividan las ramas de $y(x)$, y los puntos en los cuales la derivada cambia de signo

(porque cuando una función cambia de signo, cambia la definición del módulo de esa función).

Además se supone que el cambio de variables $y(x)$ cuyas ramas son continuas internamente. Es decir, se asume que dentro de cada rama de $y(x)$ no hay discontinuidades. De haberlas, también será necesario considerar entre los puntos que particionan el dominio de X , los puntos en los cuales $y(x)$ tenga discontinuidades.

Comenzamos a analizar uno por uno los intervalos determinados en la etapa anterior. Cada intervalo de la X hace un determinado "aporte" a la distribución de Y . Dicho aporte será hecho en forma distinta, dependiendo de si la derivada es o no igual a cero en ese intervalo.

Cuando en un intervalo la derivada es distinta de cero, podremos usar la fórmula del cambio de variables para calcular el aporte. Como vimos antes, esa fórmula involucra a la función de densidad de x , la definición de y en términos de x , y el módulo de la derivada. Como en la etapa anterior dividimos el intervalo entre otras cosas por los puntos en los cuales la derivada cambia de signo, tenemos garantizado que dentro de cada intervalo la derivada conservará su signo. Entonces tomar el módulo de la derivada se reduce a dejar la derivada tal cual está si es mayor que cero, o multiplicarla por -1 si es menor que cero.

Luego nos ocuparemos de determinar la definición de la función de densidad de X que corresponde al intervalo analizado. Planteamos el cociente dado por la fórmula, e intentamos simplificar todo lo posible. Luego hacemos la inversa de $y(x)$ para conseguir $x(y)$, y usamos esa función para dejar la expresión de la fórmula en términos de y . Dicha expresión constituirá el aporte a la distribución de Y del intervalo de X que estamos estudiando.

Sólo resta determinar a qué intervalo de Y corresponderá ese aporte. Esto se logra transformando mediante $y(x)$ los extremos x_1 y x_2 del intervalo, de modo de obtener y_1 e y_2 . Si en el intervalo analizado la derivada es negativa ($y(x)$ decreciente) entonces nos quedará $y_2 < y_1$. Llegado el caso, para evitar confusiones conviene invertir los valores de y_1 e y_2 .

Ahora bien, si en un intervalo la derivada diera cero, eso significa que en ese intervalo de X , Y no depende de X . Es decir, para ese intervalo de X , Y se mantiene constante. Entonces ese valor de Y , aunque será un punto de un continuo (el dominio de Y) tendrá probabilidad no nula, acumulará la probabilidad de la integral de la f_x para los x de ese intervalo. Es decir que si por ejemplo para $10 < x < 20$, $y=3$, entonces $P(Y=3)$ no valdrá cero, sino que

$$P(Y = 3) = P(10 < X < 20) = \int_{10}^{20} f_x \cdot dx$$

lo cual arrojará un valor mayor que cero. Entonces el aporte de este intervalo a la distribución de Y no será contribuido a $f_Y(y)$ sino a una $P_Y(y)$ que acompañará a la función de densidad.

Esto nos lleva a lo que se conoce como una distribución mixta. Por un lado tendremos la $f_Y(y)$ construida a partir de todos los intervalos de X para los cuales la

derivada no es nula. Y esa $f_Y(y)$ no cerrará a 1. Y por otro lado tendremos varios puntos con probabilidad no nula, es decir, en el caso general de varios puntos, tendremos una $P_Y(y)$. Y la integral de $f_Y(y)$, sumada a las probabilidades de todos los puntos con probabilidad no nula descriptos en $P_Y(y)$, cerrará a 1.

Cuando llegue la hora de expresar la distribución de Y , lo haremos dando tanto la $f_Y(y)$ (que tendrá los valores continuos) como la $P_Y(y)$ (que tendrá los valores discretos). Como ya se dijo antes, esto se conoce como distribución mixta, y los puntos con probabilidad no nula se denominan "puntos pesados".

Una vez determinado el aporte de cada intervalo, se procede a construir $f_Y(y)$. Cada intervalo de X en el que la derivada no es nula, aporta a $f_Y(y)$ una determinada expresión para un determinado intervalo.

Las ramas de $f_Y(y)$ estarán separadas por los puntos que determinen los intervalos de cada uno de los aportes. Y luego la definición de $f_Y(y)$ para cada rama estará dada por la suma de las expresiones de cada uno de los aportes que correspondan a intervalos en los cuales la rama esté incluida.

Si por ejemplo hubiera dos aportes:

$$y+1, \text{ para } 5 < y < 7$$

$$y^2, \text{ para } 6 < y < 8$$

entonces $f_Y(y)$ tendrá las 3 ramas siguientes:

$$y+1, \text{ para } 5 < y < 6$$

$$y+1+y^2, \text{ para } 6 < y < 7$$

$$y^2, \text{ para } 7 < y < 8$$

Nota: el método presentado no requiere hacer ningún tipo de gráfico. Sin embargo, es recomendable hacer un gráfico de $f_X(x)$ e $y(x)$ contra x , para disminuir las chances de cometer errores, sobre todo en la determinación de los puntos que dividen el dominio de X .

Problemas típicos

1) La cantidad de petróleo extraída por día de un determinado pozo, en metros cúbicos, tiene una distribución uniforme entre 200 y 300. Cada metro cúbico deja \$100 de ganancias. ¿Cómo está distribuida la probabilidad de las ganancias?

Resolución:

Si llamamos X a la cantidad de petróleo extraída, entonces:

$$f_X(x) = \begin{cases} 0,01 & 200 < x < 300 \\ 0 & \forall \text{ otro } x \end{cases}$$

y además:

$$y(x) = 100 \cdot x \quad \forall x$$

con lo cual:

$$y'(x) = 100 \quad \forall x$$

Los puntos que dividen ramas de x son el 200 y el 300, no hay puntos que dividan ramas de y , y la derivada es no nula y además conserva su signo $\forall x$.

Hay un único intervalo a estudiar:

$$200 < x < 300$$

Aplicando $y(x)$ a los extremos obtenemos que ese intervalo aporta a Y en el intervalo:

$$20000 < y < 30000$$

La densidad de X :

$$f_X(x) = 0,01$$

El módulo de la derivada:

$$\left| \frac{dy}{dx} \right| = 100$$

El aporte a Y de este intervalo de X :

$$\frac{f_X}{\left| \frac{dy}{dx} \right|} = \frac{0,01}{100} = 0,0001$$

Construimos f_Y :

$$f_Y(y) = \begin{cases} 0,0001 & 20000 < y < 30000 \\ 0 & \forall \text{ otro } y \end{cases}$$

Con lo cual obtenemos el casi obvio resultado de que las ganancias estarán distribuidas uniformemente entre 20000 y 30000.

2) Se tiene la variable aleatoria X distribuida según:

$$f_X(x) = \begin{cases} x & 0 < x < 1 \\ 2 - x & 1 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

Encuentre la distribución de $y = x^3$.

Resolución:

Tenemos $f_X(x)$ e $y(x)$. Calculamos $y'(x) = 3x^2 \quad \forall x$

Los puntos que dividen ramas de X son: 0; 1; 2. No hay puntos que dividan ramas de Y . No hay puntos en los cuales cambie el signo de la derivada. Ahora

estudiaremos uno por uno los intervalos.

- $0 < x < 1$

La derivada es mayor que cero. El módulo de la derivada es $3x$



Corresponde al intervalo: $0 < y < 1$

$$f_x(x) = x$$

$$x(y) = y^{1/3}$$

$$\left| \frac{dy}{dx} \right| = \frac{x}{3x^2} = \frac{1}{3x} = \frac{1}{3y^{1/3}}$$

- $1 < x < 2$

La derivada es mayor que cero. El módulo de la derivada es $3x^2$.

Corresponde al intervalo: $1 < y < 8$

$$f_x(x) = 2-x$$

$$x(y) = y^{1/3}$$

$$\left| \frac{dy}{dx} \right| = \frac{2-x}{3x^2} = \frac{2-y^{1/3}}{3y^{2/3}}$$

Construyendo $f_Y(y)$:

$$f_Y(y) = \left\{ \begin{array}{ll} \frac{1}{3y^{1/3}} & 0 < y < 1 \\ \frac{2-y^{1/3}}{3y^{2/3}} & 1 < y < 8 \\ 0 & \forall \text{ otro } y \end{array} \right.$$

3) En una planta embotelladora, una máquina vierte en un recipiente una determinada cantidad de líquido distribuída según:

$$f_x(x) = \begin{cases} x/4 & 0 < x < 2 \\ 4-x & 2 < x < 4 \\ \frac{4}{0} & \forall \text{ otro } x \end{cases}$$

en litros.

La capacidad del recipiente es de 3 litros, y si la máquina vierte una excesiva cantidad de líquido, la botella rebalsa.

Indique cómo está distribuida la cantidad de líquido que queda en la botella.

Resolución:

Entendamos un poco la situación. La cantidad de líquido que va a quedar en la botella es la misma cantidad de líquido que la máquina vierte, a menos que la máquina vierta más de 3 litros, en cuyo caso la cantidad en la botella sigue siendo 3 porque rebalsa. Entonces $y(x)$ nos queda así:

$$y(x) = \begin{cases} x & 0 < x < 3 \\ 3 & x > 3 \\ 0 & \forall \text{ otro } x \end{cases}$$

Con lo cual ya tenemos definidos $f_x(x)$ e $y(x)$. Calculamos $y'(x) \forall x$:

$$\frac{dy}{dx}(x) = \begin{cases} 1 & 0 < x < 3 \\ 0 & \forall \text{ otro } x \end{cases}$$

Los puntos que dividen ramas de X son: 0; 2; 4.

Los puntos que dividen ramas de Y son: 0; 3.

No hay puntos en los cuales cambie el signo de la derivada.

Entonces los puntos que dividirán los intervalos son: 0; 2; 3; 4.

Ahora estudiaremos uno por uno los intervalos:

• $0 < x < 2$

La derivada es mayor que cero. El módulo de la derivada es 1.

Corresponde al intervalo: $0 < y < 2$

$$f_x(x) = x/4$$

$$x(y) = y$$

$$\frac{f_x}{\left| \frac{dy}{dx} \right|} = \frac{x}{4} = \frac{y}{4}$$

• $2 < x < 3$

La derivada es mayor que cero. El módulo de la derivada es 1.

Corresponde al intervalo: $2 < y < 3$

$$f_x(x) = (4-x)/4$$

$$x(y) = y$$

$$\frac{f_x}{\left| \frac{dy}{dx} \right|} = \frac{4-x}{4} = \frac{4-y}{4}$$

- $3 < x < 4$

La derivada es nula. Nos encontramos frente a un punto pesado.

Y es constante y vale $Y = 3$.

$$P(Y = 3) = \int_3^4 f_X dr = \int_3^4 \frac{4-x}{x} dx = \frac{1}{8}$$

Construyendo la distribución de Y:

$$f_Y = \left\{ \begin{array}{ll} \frac{y}{4} & 0 < y < 2 \\ \frac{4-y}{4} & 2 < y < 3 \\ 0 & \forall \text{ otro } y \end{array} \right\}$$

$$P_Y = \left\{ \begin{array}{ll} \frac{1}{8} & y = 3 \\ 0 & \forall \text{ otro } y \end{array} \right\}$$

Esta distribución es mixta porque tiene una parte discreta y una parte continua. Se puede leer más sobre las distribuciones mixtas en la sección 7 del capítulo 2.

4) Se tienen la variable aleatoria X distribuida según:

$$f_X(x) = \left\{ \begin{array}{ll} x/4 & 0 < x < 2 \\ \frac{1}{4} & 2 < x < 4 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

y la variable aleatoria Y definida en términos de X según:

$$y(x) = \left\{ \begin{array}{ll} 5x & 0 < x < 1 \\ (x-3)^2 & 1 < x < 4 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

Encuentre la distribución de Y.

Resolución:

tenemos definidos $f_X(x)$ e $y(x)$. Calculamos $y'(x) \forall x$:

$$\frac{dy}{dx}(x) = \left\{ \begin{array}{ll} 5 & 0 < x < 1 \\ 2(x-3) & 1 < x < 4 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

Los puntos que dividen ramas de X son: 0; 2; 4.

Los puntos que dividen ramas de Y son: 0; 1; 4.

Advertimos que la derivada cambia de signo en $x = 3$.

Entonces los puntos que dividirán los intervalos son: 0; 1; 2; 3; 4.

Ahora estudiaremos uno por uno los intervalos:

- $0 < x < 1$

La derivada es distinta de cero y su módulo vale:

$$\left| \frac{dy}{dx} \right| = 5$$

Corresponde al intervalo: $0 < y < 5$

$$f_x(x) = x^{1/4}$$

$$x(y) = y^5$$

$$\frac{f_x}{\left| \frac{dy}{dx} \right|} = \frac{x}{4.5} = \frac{x}{20} = \frac{y}{100}$$

• $1 < x < 2$

La derivada es distinta de cero.

Corresponde al intervalo: $1 < y < 4$

El módulo de la derivada es $|2(x-3)|$. Como en este intervalo x va entre 1 y 2, $x < 3$ con lo cual el paréntesis va a dar negativo. Por lo tanto para hallar el módulo multiplicamos por -1 y queda:

$$\left| \frac{dy}{dx} \right| = 2(3 - x)$$

$$f_x(x) = x^{1/4}$$

$$y = (x-3)^2 \Rightarrow \sqrt{y} = |x-3|$$

como $x < 3$, queda:

$$\sqrt{y} = 3 - x \Rightarrow x(y) = 3 - \sqrt{y}$$

$$\frac{f_x}{\left| \frac{dy}{dx} \right|} = \frac{x}{4 \cdot 2 \cdot (3-x)} = \frac{x}{8 \cdot (3-x)} = \frac{3 - \sqrt{y}}{8\sqrt{y}}$$

• $2 < x < 3$

La derivada es distinta de cero.

Corresponde al intervalo: $0 < y < 1$

El módulo de la derivada es $|2(x-3)|$. Como en este intervalo x va entre 2 y 3, $x < 3$ con lo cual el paréntesis va a dar negativo. Por lo tanto para hallar el módulo multiplicamos por -1 y queda:

$$\left| \frac{dy}{dx} \right| = 2(3 - x)$$

$$f_x(x) = x^{1/4}$$

$$y = (x-3)^2 \Rightarrow \sqrt{y} = |x-3|$$

como $x < 3$, queda:

$$\sqrt{y} = 3 - x \Rightarrow x(y) = 3 - \sqrt{y}$$

$$\frac{f_x}{\left| \frac{dy}{dx} \right|} = \frac{1}{4 \cdot 2 \cdot (3-x)} = \frac{1}{8 \cdot (3-x)} = \frac{1}{8\sqrt{y}}$$

- $3 < x < 4$

La derivada es distinta de cero.

Corresponde al intervalo: $0 < y < 1$

El módulo de la derivada es $|2(x-3)|$. Como en este intervalo x va entre 3 y 4, $x > 3$ con lo cual el paréntesis va a dar positivo. Queda:

$$\left| \frac{dy}{dx} \right| = 2(x-3)$$

$$f_x(x) = 1/4$$

$$y = (x-3)^2 \Rightarrow \sqrt{y} = |x-3|$$

como $x > 3$, queda:

$$\sqrt{y} = x-3 \Rightarrow x(y) = 3 + \sqrt{y}$$

$$\frac{f_x}{\left| \frac{dy}{dx} \right|} = \frac{1}{4 \cdot 2 \cdot (x-3)} = \frac{1}{8 \cdot (x-3)} = \frac{1}{8\sqrt{y}}$$

Resumiendo, los aportes son:

$$\frac{y}{100} \quad 0 < y < 5$$

$$\frac{3 - \sqrt{y}}{8\sqrt{y}} \quad 1 < y < 4$$

$$\frac{1}{8\sqrt{y}} \quad 0 < y < 1$$

$$\frac{1}{8\sqrt{y}} \quad 0 < y < 1$$

Construimos $f_Y(y)$:

$$f_Y(y) = \left\{ \begin{array}{ll} \frac{y}{100} + \frac{1}{8\sqrt{y}} + \frac{1}{8\sqrt{y}} & 0 < y < 1 \\ \frac{y}{100} + \frac{3 - \sqrt{y}}{8\sqrt{y}} & 1 < y < 4 \\ \frac{y}{100} & 4 < y < 5 \\ 0 & \forall \text{ otro } y \end{array} \right.$$



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/processtext/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 4 de mayo de 2004

Media o Esperanza o Valor Esperado

Dada una variable aleatoria, nos puede interesar tener una idea de qué valor podríamos esperar que asuma si se hace el experimento al cual está asociada. Por ejemplo, nos puede interesar calcular el consumo medio por hora de una máquina, la cantidad de clientes que podemos esperar tener en un determinado día, o la cantidad media de líquido que la embotelladora envasa en las botellas. Para ese fin utilizamos la media o esperanza matemática.

Dada X una variable aleatoria, si su esperanza $E(X)$ existe, vale:

$$E(X) = \sum_{-\infty}^{+\infty} x P_X(x)$$

si X es discreta

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

si X es continua

Como $E(X)$ está definida a partir de una sumatoria o integral, resulta ser un operador lineal, con lo cual se puede demostrar fácilmente que:

$$E(aX + b) = E(aX) + E(b) = aE(X) + b \quad \text{con } a, b \in \mathfrak{R}$$

de donde también se observa que la esperanza de una constante es la propia constante.

En el caso general, en vez de interesarnos calcular la esperanza de X , nos puede interesar calcular la esperanza de una función $\varphi(X)$. Si $Y = \varphi(X)$, vale:

$$E(Y) = \sum_{-\infty}^{+\infty} \varphi(x) P_X(x)$$

si X es discreta

$$E(Y) = \int_{-\infty}^{+\infty} \varphi(x) f_X(x) dx$$

si X es continua

Comentarios

- 1) Podemos pensar en la media como el valor que obtendríamos si tomáramos infinitas muestras de una variable aleatoria e hiciéramos el promedio de sus valores.
- 2) La media no tiene necesariamente que ser un valor posible.

Ejemplos

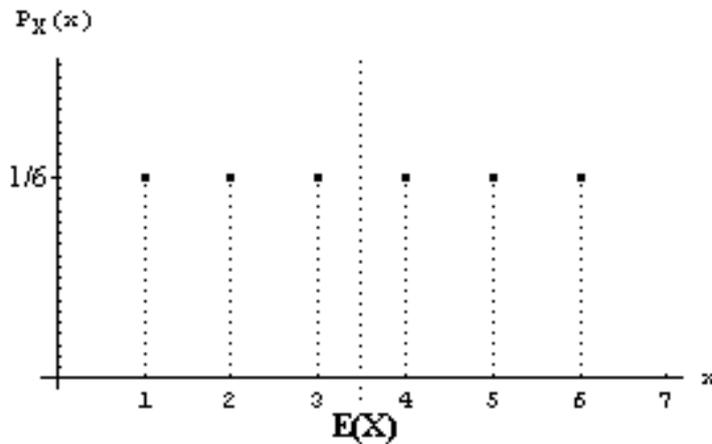
$$P_X(x) = \begin{cases} 1/3 & x=1 \\ 1/3 & x=2 \\ 1/3 & x=3 \\ 0 & \forall \text{ otro } x \end{cases}$$

1) Sea X discreta distribuida según

$$E(X) = \sum_{-\infty}^{\infty} x P_X(x) = \sum_{x=1}^3 x \frac{1}{3} = \frac{1}{3} \sum_{x=1}^3 x = \frac{6}{3} = 2$$

Como era de esperar, si X podía ser al azar 1, 2 ó 3, la media es 2.

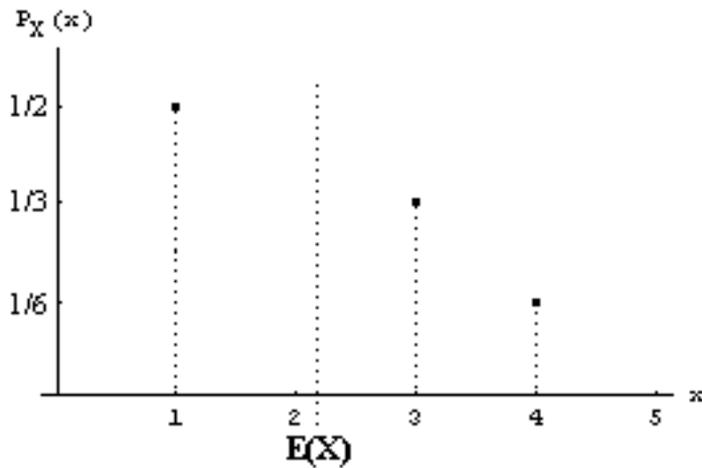
2) Veamos ahora el ejemplo del dado:



$$P_X(x) = \begin{cases} 1/6 & x=1 \\ 1/6 & x=2 \\ 1/6 & x=3 \\ 1/6 & x=4 \\ 1/6 & x=5 \\ 1/6 & x=6 \\ 0 & \forall \text{ otro } x \end{cases}$$

$$E(X) = \sum_{-\infty}^{\infty} x P_X(x) = \sum_{x=1}^6 x \frac{1}{6} = \frac{1}{6} \sum_{x=1}^6 x = \frac{21}{6} = 3.5$$

Este ejemplo nos muestra que la media no tiene por qué necesariamente ser un valor posible. Es solamente el valor ESPERADO matemáticamente de la distribución. Como se dijo antes, podemos imaginarlo como el valor que obtendríamos si tomáramos infinitas muestras de la variable aleatoria e hiciéramos el promedio de sus valores.



$$P_X(x) = \begin{cases} 1/2 & x=1 \\ 1/3 & x=3 \\ 1/6 & x=4 \\ 0 & \forall \text{ otro } x \end{cases}$$

3) Sea X discreta distribuida según

$$E(X) = \sum_{-\infty}^{\infty} x P_X(x) = \sum_{x=1}^{\infty} x P_X(x) =$$

$$= 1 \cdot \frac{1}{2} + 3 \cdot \frac{1}{3} + 4 \cdot \frac{1}{6} = 2.1667$$

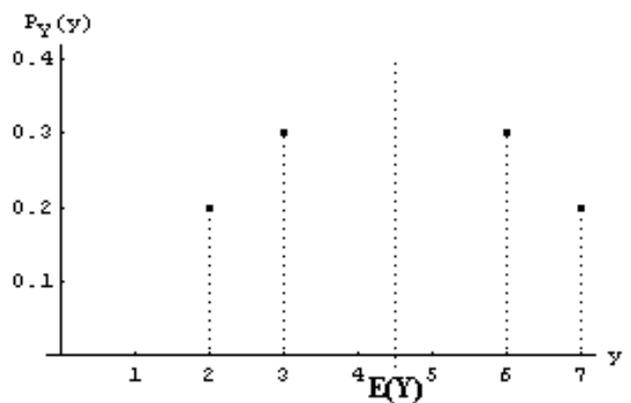
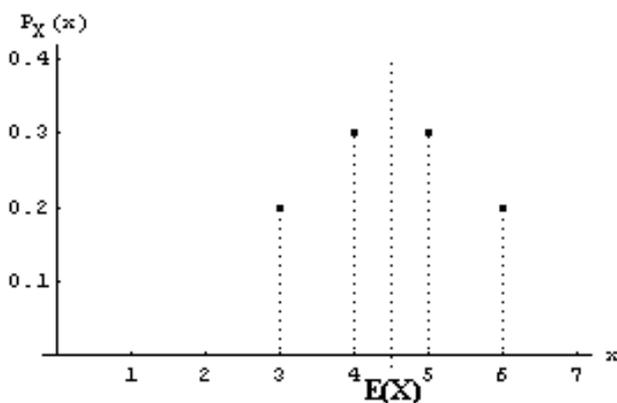
Este ejemplo nos muestra que la media de una distribución tampoco es necesariamente el valor más probable.

4) Sean X e Y distribuidas según:

$$P_X(x) = \begin{cases} 0,2 & x=3 \\ 0,3 & x=4 \\ 0,3 & x=5 \\ 0,2 & x=6 \\ 0 & \forall \text{ otro } x \end{cases} \quad P_Y(y) = \begin{cases} 0,2 & x=2 \\ 0,3 & x=3 \\ 0,3 & x=6 \\ 0,2 & x=7 \\ 0 & \forall \text{ otro } x \end{cases}$$

$$E(X) = \sum_{-\infty}^{\infty} x P_X(x) = 3 \cdot 0,2 + 4 \cdot 0,3 + 5 \cdot 0,3 + 6 \cdot 0,2 = 4,5$$

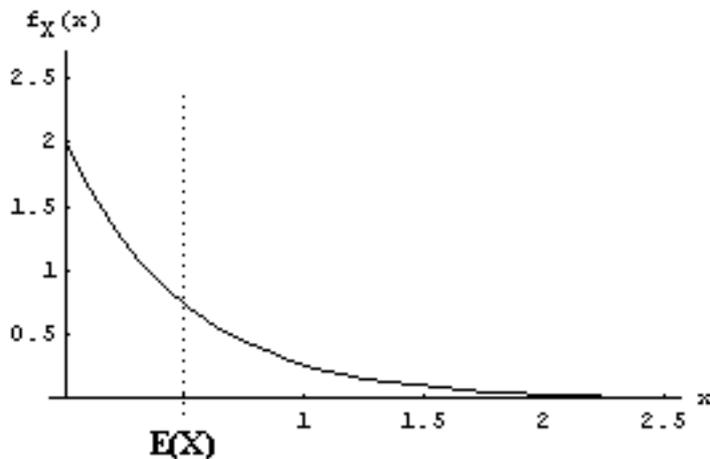
$$E(Y) = \sum_{-\infty}^{\infty} y P_Y(y) = 2 \cdot 0,2 + 3 \cdot 0,3 + 6 \cdot 0,3 + 7 \cdot 0,2 = 4,5$$



Vemos que si la distribución es simétrica, la esperanza "no se entera" de si los valores con probabilidad no nula están más cercanos o más espaciados.

5) La media de las distribuciones continuas es análoga a la de las distribuciones discretas, y en general cumple las mismas propiedades.

- No necesariamente coincide con el valor más probable



- No necesariamente es un valor posible
- Si la distribución es simétrica, no se entera de si los valores con probabilidad no nula están más cercanos o más espaciados.

A modo de ejemplo tomemos:

$$f_X(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{+\infty} x 2e^{-2x} dx = \frac{1}{2}$$

6) La media puede no existir.

$$f_X(x) = \frac{1}{\pi(1+x^2)} \quad \forall x \in \mathfrak{R}$$

Tomemos por ejemplo la función de densidad:

En ningún momento es negativa y su integral da 1, con lo cual es efectivamente una función de densidad.

Calculemos la media de esta distribución:

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{-\infty}^{+\infty} \frac{x}{\pi(1+x^2)} dx = \frac{\ln(1+x^2)}{2\pi} \Big|_{-\infty}^{+\infty} = \infty - \infty = ?$$

Vemos que esta integral no existe, por lo tanto la esperanza de esta distribución no existe.

Moda o modo o valor más probable

Es otra característica que podemos calcular de una distribución. La moda de una distribución es el valor más probable. Es decir, si X es una variable aleatoria

discreta, $Mo(X)$ es el x (o los x) tales que $P_x(x)$ es máxima. Si X es continua, es el o los x tales que $f_x(x)$ es máxima.

En el ejemplo n° 3 de la media, la moda es 1.

En el ejemplo n° 5 de la media, la moda es 0.

- La moda de una distribución no necesariamente es única.

En el ejemplo n° 2 de la media, los 6 valores posibles son la moda.

- A diferencia de la media, la moda es necesariamente un valor posible.

- Como caso particular, la media y la moda pueden coincidir, por ejemplo en:

$$P_X(x) = \begin{cases} 0,3 & x=1 \\ 0,4 & x=2 \\ 0,3 & x=3 \\ 0 & \forall \text{ otro } x \end{cases}$$

donde $E(X) = Mo(X) = 2$

Mediana

Definimos la mediana de una distribución continua de X como M tal que $P(X < M) = P(X > M) = 0,5$. Es decir, es el valor que se encuentra en el medio, probabilísticamente hablando.

Tomando una de las dos formas, hallamos M resolviendo:

$$\int_{-\infty}^M f_X(x) dx = 0,5 \quad \text{o bien} \quad \int_M^{+\infty} f_X(x) dx = 0,5 \quad \text{según convenga.}$$

- A diferencia de la media, la mediana siempre existe, y además es menos sensible a las distribuciones que están espaciadas hacia uno de sus lados (como el ejemplo n° 5 de la media).

- La mediana no necesariamente es única (como en cualquier análogo continuo del ejemplo n° 4 de la media, en el cual habría 2 medianas).

Ejemplo:

$$f_X(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Hallar la mediana de

$$\int_{-\infty}^M f_X(x) dx = 0,5 \Rightarrow \int_0^M 2e^{-2x} dx = -e^{-2x} \Big|_0^M = 1 - e^{-2M} = 0,5 \Rightarrow M = \frac{-\ln(0,5)}{2} = \frac{\ln(2)}{2} = 0,35$$

Problemas típicos

1) Halle la media y la moda de X, donde X está distribuida según:

$$P_X(x) = \left\{ \begin{array}{ll} 0,4 & x = -1 \\ 0,1 & x = 1 \\ 0,3 & x = 2 \\ 0,2 & x = 3 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

Resolución:

$$E(X) = \sum_{-\infty}^{+\infty} x P_X(x) = (-1) \cdot 0,4 + 1 \cdot 0,1 + 2 \cdot 0,3 + 3 \cdot 0,2 = 0,9$$

Mo(X) = valor más probable = -1

2) La longitud de las varillas fabricadas por una máquina es la variable aleatoria X distribuida según:

$$f_X(x) = \left\{ \begin{array}{ll} x^2 & 0 \leq x \leq 1 \\ \frac{1}{3} & 1 \leq x \leq 3 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

¿Cuál es la longitud media de las varillas?

Resolución:

"La longitud media de las varillas" se refiere a "La media de la longitud de las varillas", es decir, la media de X.

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^1 x x^2 dx + \int_1^3 x \frac{1}{3} dx = 1,583$$

3) Con la misma X del ejercicio anterior, encuentre la esperanza de:

a) Y = 3X-5

b) Z = X²+2X

Resolución:

a) $E(aX + b) = E(aX) + E(b) = aE(X) + b$

$E(3X - 5) = 3E(X) - 5$

Y habíamos calculado que $E(X) = 1.583$, con lo cual:

$E(Y) = -0,25$

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

b) Tenemos 2 formas de resolverlo, usando

Una forma consiste en tomar $h(x) = x^2 + 2x$ y hacer:

$$E(h(X)) = \int_{-\infty}^{+\infty} h(x) f_X(x) dx = \int_0^1 (x^2 + 2x) x^2 dx + \int_1^2 (x^2 + 2x) \frac{1}{3} dx = 6,256$$

Pero hay otra forma de hacerlo con menos cuentas. Aprovechando la linealidad del operador esperanza, hacemos:

$$E(x^2 + 2x) = E(x^2) + 2.E(x)$$

con lo cual podemos calcular la esperanza de x^2 en vez de la de $x^2 + 2x$

Tomamos $h(x) = x^2$

$$E(x^2) = \int_{-\infty}^{+\infty} h(x) f_X(x) dx = \int_0^1 x^2 x^2 dx + \int_1^2 x^2 \frac{1}{3} dx = 3,089$$

Y luego $E(Z) = 3,089 + 2.1,583 = 6,256$



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/process/text/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 4 de mayo de 2004_

Varianza

Vimos que la media o esperanza nos da una idea de qué valor podríamos esperar que asuma una determinada variable aleatoria, si se lleva a cabo el experimento al cual está asociada. Es decir, la media es una **medida de posición**.

Asimismo, vimos que la esperanza no nos proporciona información acerca de si los valores que puede tomar la variable aleatoria se encuentran cercanos o espaciados. Por eso utilizaremos otra herramienta matemática denominada varianza. La varianza es una medida de cuánto tienden los valores de una variable aleatoria a alejarse de la media de la misma. La varianza es una **medida de dispersión**.

Dada X una variable aleatoria, si su varianza σ_x^2 existe, vale:

$$\text{Var}(X) = \sigma_x^2 = E((X - E(X))^2) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f_x(x) dx$$

Vemos que la varianza es la esperanza de los cuadrados de las distancias entre los valores de la variable y el valor medio de la distribución. Si los valores de X están muy dispersos, los $E(X) - X$ tenderán a ser más grandes, y la varianza tiende a ser mayor. Observamos también que como las diferencias están al cuadrado, no importa si son positivas (X a la derecha de la media) o negativas (X a la izquierda de la media). O sea que todas "suman".

Operando con la fórmula de arriba se llega a otra fórmula para la varianza, que a menudo resulta más práctica:

$$\sigma_x^2 = E(X^2) - E(X)^2$$

La varianza también presenta la siguiente propiedad:

$$\sigma^2(aX + b) = a^2\sigma_x^2 \quad \text{con } a, b \in \mathfrak{R}$$

Mas adelante en esta misma sección se demuestran las fórmulas y propiedades.

Comentarios

- La varianza es una medida de cuánto tienden los valores de una variable aleatoria a alejarse de la media de la misma. Es decir que si la varianza es chica, la distribución se encuentra concentrada alrededor de la media, y si es grande, se encuentra más esparcida, más dispersa.
- Como la varianza se define a partir de la media, puede, al igual que esta, no existir.

Ejemplo

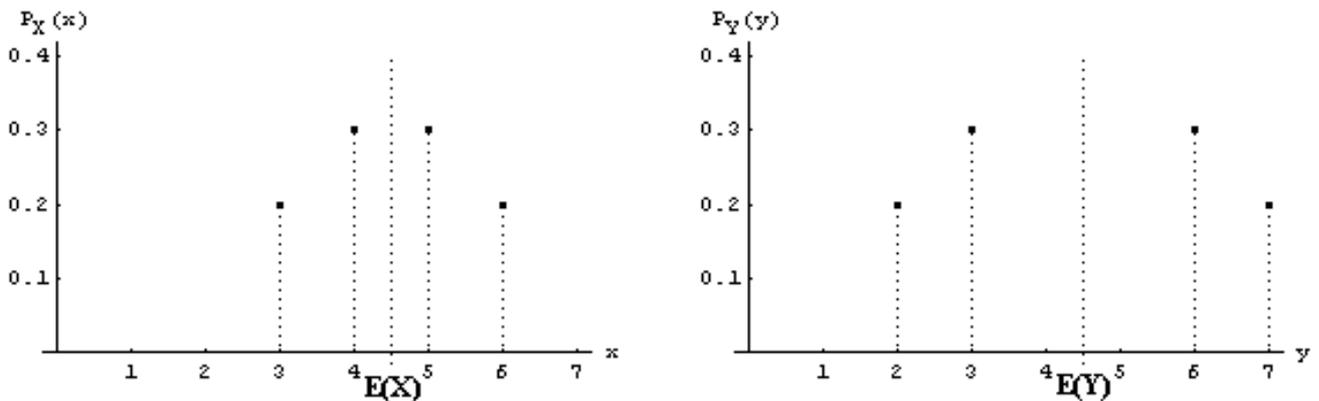
Volvamos al ejemplo 5 de la media:

$$P_X(x) = \left. \begin{array}{l} 0,2 \quad x=3 \\ 0,3 \quad x=4 \\ 0,3 \quad x=5 \\ 0,2 \quad x=6 \\ 0 \quad \forall \text{ otro } x \end{array} \right\}$$

Tenemos la distribución de X e Y, y calculamos sus medias:

$$P_Y(y) = \left. \begin{array}{l} 0,2 \quad x=2 \\ 0,3 \quad x=3 \\ 0,3 \quad x=6 \\ 0,2 \quad x=7 \\ 0 \quad \forall \text{ otro } x \end{array} \right\} \begin{array}{l} E(X) = \sum_{-\infty}^{\infty} x P_X(x) = 3 \cdot 0,2 + 4 \cdot 0,3 + 5 \cdot 0,3 + 6 \cdot 0,2 = 4,5 \\ E(Y) = \sum_{-\infty}^{\infty} y P_Y(y) = 2 \cdot 0,2 + 3 \cdot 0,3 + 6 \cdot 0,3 + 7 \cdot 0,2 = 4,5 \end{array}$$

Habíamos observado que las medias de X e Y son iguales, a pesar de que Y está más dispersa que X:



Veamos qué sucede con las varianzas:

$$E(X^2) = \sum_{-\infty}^{\infty} x^2 P_X(x) = 3^2 \cdot 0,2 + 4^2 \cdot 0,3 + 5^2 \cdot 0,3 + 6^2 \cdot 0,2 = 21,3$$

$$E(Y^2) = \sum_{-\infty}^{\infty} y^2 P_Y(y) = 2^2 \cdot 0,2 + 3^2 \cdot 0,3 + 6^2 \cdot 0,3 + 7^2 \cdot 0,2 = 24,1$$

$$\sigma_x^2 = E(X^2) - E(X)^2 = 21,3 - 4,5^2 = 1,05$$

$$\sigma_y^2 = E(Y^2) - E(Y)^2 = 24,1 - 4,5^2 = 3,85$$

Vemos que la varianza de Y es casi 4 veces mayor que la varianza de X. Esto refleja que las probabilidades de los valores de Y se encuentran más alejados

de la media que los de X.

Desvío estándar

El desvío estándar σ_x de una variable aleatoria X se define como la raíz cuadrada positiva de su varianza.

$$\sigma_x = \sqrt{\sigma_x^2}$$

Unidades

Si X es la longitud de los tornillos fabricados por una máquina, entonces las unidades de X podrían ser, por ejemplo, cm.

A su vez, como la media o esperanza es el valor esperado de X, tiene la misma forma que X (sea un valor posible realmente o no). Entonces las unidades de E(X) deben ser las mismas que las de X, es decir, cm.

La varianza se puede obtener, por ejemplo, $\sigma_x^2 = E(X^2) - E(X)^2$, donde se ve claramente que las unidades de la varianza son cm^2 .

Y como el desvío estándar se define como la raíz cuadrada de la varianza, entonces sus unidades vuelven a ser las de X, es decir, cm.

Demostraciones

Comenzaremos por probar que:

$$\sigma_x^2 = E((X - E(X))^2) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f_X(x) dx = E(X^2) - E(X)^2$$

Partimos de:

$$\sigma_x^2 = E((X - E(X))^2)$$

Como dada una distribución, su esperanza es una constante, vamos a escribir, por claridad, μ_x en vez de E(X).

$$\sigma_x^2 = E((X - \mu_x)^2)$$

Notemos que $(X - \mu_x)^2$ es una función de X. Luego su esperanza vale:

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 f_X(x) dx$$

Con lo cual llegamos a la segunda fórmula dada. Ahora desarrollemos el cuadrado:

$$\int_{-\infty}^{\infty} (x^2 + \mu_x^2 - 2x\mu_x) f_X(x) dx$$

Abrimos la integral en tres:

$$\int_{-\infty}^{\infty} x^2 f_X(x) dx + \int_{-\infty}^{\infty} \mu_x^2 f_X(x) dx - \int_{-\infty}^{\infty} 2x\mu_x f_X(x) dx$$

Como 2 y μ_x son constantes, salen de las integrales:

$$\int_{-\infty}^{\infty} x^2 f_X(x) dx + \mu_x^2 \int_{-\infty}^{\infty} f_X(x) dx - 2\mu_x \int_{-\infty}^{\infty} x f_X(x) dx$$

El primer término es, por definición de esperanza de una función, $E(X^2)$. En el segundo término, la integral da uno. La integral del tercer término es por definición la esperanza de X, es decir, μ_x . Queda:

$$E(X^2) + \mu_x^2 - 2\mu_x^2$$

Con lo cual llegamos a la tercera fórmula dada:

$$\sigma_x^2 = E(X^2) - E(X)^2$$

Ahora vamos a demostrar la propiedad:

$$\sigma^2(aX + b) = a^2\sigma_x^2$$

Llamaremos $Y = aX + b$. Luego por definición de varianza:

$$\sigma_y^2 = E((Y - E(Y))^2)$$

Reemplazando Y por $aX + b$ obtenemos:

$$\sigma_{aX+b}^2 = E((aX + b - E(aX + b))^2) = E((aX + b - aE(X) - b)^2) = E((aX - a\mu_x)^2)$$

Sacando factor común a, y sacándola del cuadrado y de la esperanza, queda:

$$\sigma_{aX+b}^2 = E(a^2(X - \mu_x)^2) = a^2 E((X - \mu_x)^2)$$

El segundo factor es por definición la varianza de X. Luego, como queríamos demostrar:

$$\sigma^2(aX + b) = a^2\sigma_x^2$$

Puede parecer extraño que la b no aparezca en la varianza de $aX + b$, pero no lo es.

La constante b no tiene ninguna influencia en la varianza porque es una constante que aparece sumando, y que a lo sumo puede correr la distribución hacia la izquierda o hacia la derecha, es decir, cambiar la **posición**, pero no la **dispersión**.

Además podemos hacer el comentario de que la varianza de una constante es cero, porque la varianza es una medida de dispersión, y como una constante es un punto, no tiene dispersión. Luego su varianza es cero.

Problemas típicos

1) Halle varianza y el desvío estándar de X, donde X está distribuida según:

$$P_X(x) = \begin{cases} 0,4 & x = -1 \\ 0,1 & x = 1 \\ 0,3 & x = 2 \\ 0,2 & x = 3 \\ 0 & \forall \text{ otro } x \end{cases}$$

Resolución:

$$E(X) = \sum_{-\infty}^{+\infty} x P_X(x) = (-1) \cdot 0,4 + 1 \cdot 0,1 + 2 \cdot 0,3 + 3 \cdot 0,2 = 0,9$$

$$E(X^2) = \sum_{-\infty}^{+\infty} x^2 P_X(x) = (-1)^2 \cdot 0,4 + 1^2 \cdot 0,1 + 2^2 \cdot 0,3 + 3^2 \cdot 0,2 = 3,5$$

$$\sigma_x^2 = E(X^2) - E(X)^2 = 2,69$$

$$\sigma_x = \sqrt{\sigma_x^2} = 1,64$$

2) La longitud en cm. de las varillas fabricadas por una máquina es la variable aleatoria X distribuida según:

$$f_X(x) = \begin{cases} x^2 & 0 \leq x \leq 1 \\ \frac{1}{3} & 1 \leq x \leq 3 \\ 0 & \forall \text{ otro } x \end{cases}$$

a) ¿Cuál es la varianza de la longitud media de las varillas?

b) Si a las varillas se las corta a la mitad y se les agrega una punta de 1 cm., ¿Cuál es la varianza de la longitud de las nuevas varillas?

Resolución:

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^1 x x^2 dx + \int_1^3 x \frac{1}{3} dx = 1,583$$

a)

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_0^1 x^2 x^2 dx + \int_1^3 x^2 \frac{1}{3} dx = 3,089$$

$$\sigma_x^2 = E(X^2) - E(X)^2 = 0,582$$

$$\sigma^2(aX + b) = a^2 \sigma_x^2 \Rightarrow \sigma^2\left(\frac{1}{2}x + 1\right) = \left(\frac{1}{2}\right)^2 \sigma_x^2 = 0,145$$

b)

3) Si dos máquinas producen piezas cuyas longitudes son variables aleatorias de igual media, pero la varianza de la longitud de las piezas fabricadas por la máquina A es mayor que la varianza de las de B, y es importante que todas las piezas sean lo más parecidas posibles, ¿cuál máquina decidiría comprar?

Resolución:

Como la varianza es una medida de la tendencia de los valores de la variable a alejarse de la media, eligiendo la máquina B las piezas fabricadas tenderán a ser de longitudes más parecidas.



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/processtext/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 4 de mayo de 2004_

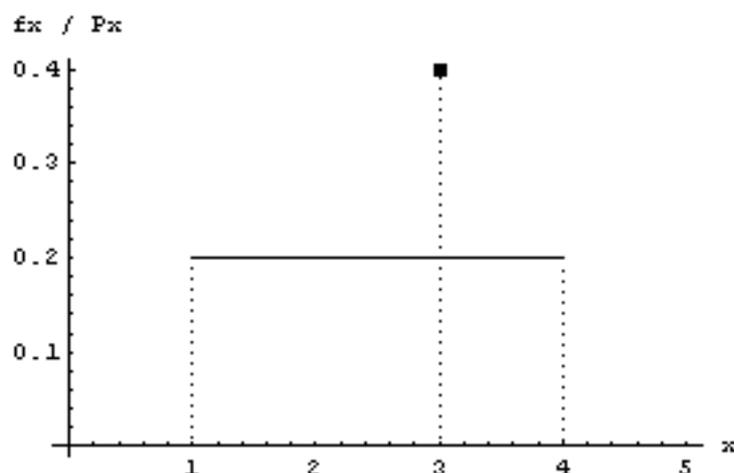
Variables aleatorias mixtas

Hasta ahora vimos que las variables aleatorias pueden ser discretas o continuas.

Hay un tercer tipo híbrido de variable aleatoria: la mixta.

Podemos pensar una variable aleatoria mixta como una continua pero que tiene valores no nulos para uno o más puntos. Dichos puntos se denominan puntos pesados.

Por ejemplo:



$$\left[\begin{array}{l} f_x(x) = \begin{cases} 1/5 & 1 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases} \\ P_x(x) = \begin{cases} 2/5 & x = 3 \\ 0 & \forall \text{ otro } x \end{cases} \end{array} \right]$$

La distribución es lo que se encuentra entre las llaves grandes, es decir, el agrupamiento de la semidistribución discreta y la semidistribución continua

También se lo puede escribir así:

$$f_x(x) = \begin{cases} 1/5 & 1 < x < 4 \\ 2/5 & x = 3 \\ 0 & \forall \text{ otro } x \end{cases}$$

En esta notación, las ramas discretas se distinguen por ser las que contienen igualdades ($x=3$) y las continuas por ser las que tienen intervalos ($1 < x < 4$). Pero no usaremos esta notación porque estamos siguiendo la convención de usar la letra f para referirnos a distribuciones continuas puras. Sin embargo muchos autores usan la letra f tanto para las distribuciones continuas como también para las distribuciones discretas y mixtas.

Función de distribución

Hallaremos la función de distribución de la variable mixta que dimos como ejemplo. Recordemos que el proceso de construir la función de distribución acumulada puede pensarse como que se recorre el dominio de la variable desde menos infinito hacia la derecha y se van recolectando probabilidades o masas de probabilidades.

Hasta el 1 no hay nada acumulado, por lo tanto para $-\infty < x < 1$ la $F(x)$ vale cero. A partir del 1 se comienza a acumular la densidad $1/5$, y se sigue hasta el 3.

Entonces la expresión de la $F(x)$ entre el 1 y el 3 es de la forma $x/5 + C$, donde C debe ser tal que la función resulte continua en el 1 porque en todo punto que no sea pesado la F debe ser continua, como ocurre como las variables continuas, que como no tienen puntos pesados, la F siempre es continua.

Entonces para $x = 1$, $x/5 + C$ debe valer cero. Luego $C = -1/5$.

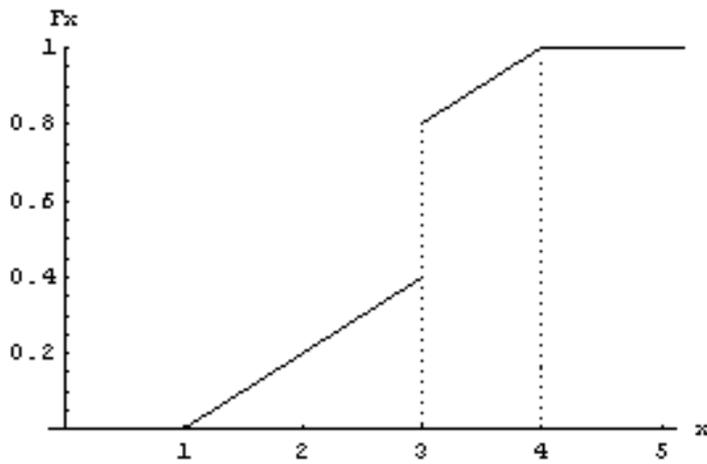
Entonces para $1 < x < 3$, $F(x) = x/5 - 1/5 = (x-1)/5$

En $x = 3$ la $F(x)$ tiene un salto, porque se está acumulando la probabilidad no nula del punto pesado. La rama que llega hasta el 3, en el 3 vale $(3-1)/5 = 2/5$. Le sumamos la probabilidad del punto pesado, $2/5$, y entonces la siguiente rama empezará en la altura $4/5$. Como la densidad que vamos a acumular a partir del 3 es la misma que antes, la expresión de la F será también $x/5 + C$, pero ahora C tendrá otro valor porque está "levantada" con respecto a la rama anterior por el salto que hubo. Dijimos que en $x=3$ debe valer $4/5$. Entonces $3/5 + C = 4/5$. Luego $C = 1/5$.

Entonces para $3 < x < 4$, $F(x) = x/5 + 1/5 = (x+1)/5$

A partir del 4, la F vale 1

La $F_x(x)$ queda:



$$F_X(x) = \begin{cases} 0 & x < 1 \\ \frac{x-1}{5} & 1 < x < 3 \\ \frac{5}{x+1} & 3 < x < 4 \\ \frac{5}{1} & x > 4 \end{cases}$$

Vemos que la función de distribución de una variable mixta se asemeja a la de una variable continua, pero en vez de ser continua como la de ésta, es continua a trozos. Los puntos de discontinuidad son los puntos pesados de la variable. Además en dichos puntos no existe el límite desde la izquierda, pero sí desde la derecha. Esto se debe a que la definición de la F es con un menor o igual (\leq).

Esperanza y varianza

Para calcular la esperanza de una variable mixta se aplica la definición de esperanza para variables discretas para las ramas discretas, y la para continuas para las ramas continuas, y luego se suman los dos resultados.

Para la distribución del ejemplo:

$\sum_{-\infty}^{+\infty} x P_X(x)$
 involucra una sola rama ($x=3$). Luego la parte discreta de la esperanza vale: $3 \cdot \frac{2}{5} = \frac{6}{5}$.

$\int_{-\infty}^{+\infty} x f_X(x) dx$
 involucra también una sola rama ($1 < x < 4$). Luego la parte continua de la esperanza es $\int_1^4 \frac{x}{5} dx = \frac{3}{2}$

esperanza es ¹

Luego $E(X) = 2.7$, lo cual es bastante coherente porque da cercano a la masa de probabilidad continua, pero desviado hacia el lado del punto pesado.

Con respecto a la varianza, no hay nada nuevo porque ésta se calcula a partir de la esperanza. Para la variable del ejemplo:

$$\int_1^3 \frac{x^2}{5} dx = \frac{18}{5} + \frac{21}{5} = \frac{39}{5} = 7.8$$

$$E(X^2) = 3^2 \cdot \frac{2}{5} + \frac{39}{5} = 7.8$$

$$E(X)^2 = 2.7^2 = 7.29$$

$$\text{Luego } \sigma_x^2 = 7.8 - 7.29 = 0.51$$

Problemas típicos

1) Se tiene una variable aleatoria con la siguiente distribución:

$$f_x(x) = \begin{cases} \frac{x}{8} & 0 < x < 2 \\ \frac{4-x}{8} & 2 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

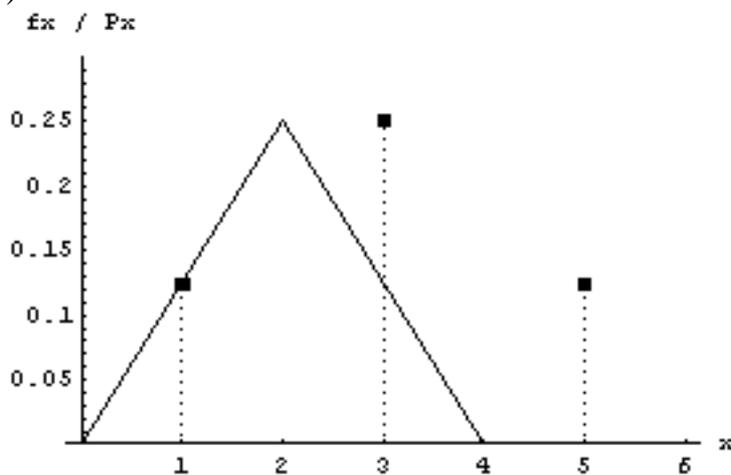
$$P_x(x) = \begin{cases} 1/8 & x=1 \\ 1/4 & x=3 \\ 1/8 & x=5 \\ 0 & \forall \text{ otro } x \end{cases}$$

Se pide:

- Grafique la función de densidad
- Calcule $P(X < 2)$ y $P(X < 3)$
- Construya y grafique $F_x(x)$
- Calcule la media y la varianza

Resolución:

a)



- b) Hasta $x=2$ hay acumulada la mitad del triángulo ($1/4$) más el punto pesado $x=1$, es decir $P(X < 2) = 1/4 + 1/8 = 3/8$
 Hasta $x=3$ hay acumulada un área adicional de $3/16$. El punto pesado de $x=3$ no lo

contamos porque nos piden $P(X < 3)$. Si nos pidieran $P(X \leq 3)$, sí lo contaríamos. Entonces $P(X < 3)$ es el $3/8$ que teníamos hasta $x=2$, más el área entre el 2 y el 3, que vale $3/16$.

Luego, $P(X < 3) = 9/16$

c)

$$F_x(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{16} & 0 < x < 1 \\ \frac{x^2}{16} + \frac{2}{16} & 1 < x < 2 \\ \frac{x}{2} - \frac{x^2}{16} - \frac{3}{8} & 2 < x < 3 \\ \frac{x}{2} - \frac{x^2}{16} - \frac{2}{16} & 3 < x < 4 \\ \frac{7}{8} & 4 < x < 5 \\ 1 & x > 5 \end{cases}$$

d) Sumamos las esperanzas de la parte continua y la parte discreta:

$$E(X) = \int_0^1 \frac{x^2}{8} dx + \int_2^4 \frac{x(4-x)}{8} dx + 1 \cdot \frac{1}{8} + 3 \cdot \frac{1}{4} + 5 \cdot \frac{1}{8} = \frac{1}{3} + \frac{2}{3} + \frac{1}{8} + \frac{3}{4} + \frac{5}{8} = 2,5$$

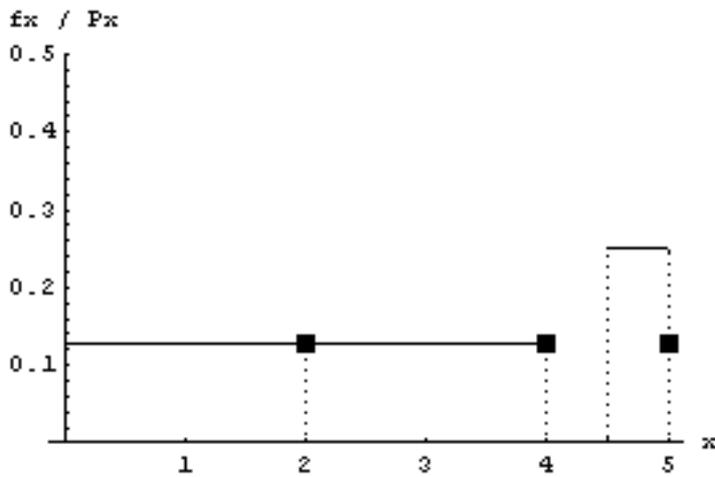
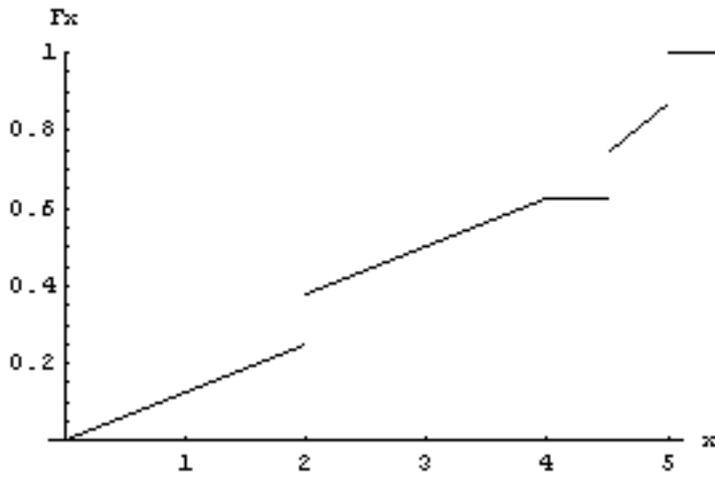
La varianza se puede calcular igual que en el ejemplo.

2) Dada la siguiente función de distribución, halle la distribución de la variable.

$$F_x(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{8} & 0 < x < 2 \\ \frac{x}{8} + \frac{1}{8} & 2 < x < 4 \\ \frac{5}{8} & 4 < x < 4,5 \\ \frac{x}{4} - \frac{3}{8} & 4,5 < x < 5 \\ 1 & x > 5 \end{cases}$$

Resolución:

Primero graficamos la función de distribución para entender lo que estamos haciendo. Luego construimos la parte continua de la distribución, simplemente derivando rama a rama la función de distribución.



Queda:

$$f_x(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{8} & 0 < x < 2 \\ \frac{1}{8} & 2 < x < 4 \\ 0 & 4 < x < 4,5 \\ \frac{1}{4} & 4,5 < x < 5 \\ 0 & x > 5 \end{cases}$$

Arreglándola un poco:

$$f_x(x) = \begin{cases} \frac{1}{8} & 0 < x < 4 \\ \frac{1}{4} & 4,5 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

Luego para la parte discreta, miramos los saltos de la función de distribución:

Salta 1/8 en $x = 2$

Salta 1/8 en $x = 4$

Salta 1/8 en $x = 5$

Luego ya podemos construir la distribución:

$$\left. \begin{aligned} f_x(x) &= \begin{cases} \frac{1}{8} & 0 < x < 4 \\ \frac{1}{4} & 4,5 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases} \\ P_x(x) &= \begin{cases} \frac{1}{8} & x = 2 \\ \frac{1}{8} & x = 4 \\ \frac{1}{8} & x = 5 \\ 0 & \forall \text{ otro } x \end{cases} \end{aligned} \right\}$$



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/process/text/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 4 de mayo de 2004_

Distribución condicionada o truncada

"La longitud de las varillas fabricadas por una máquina es una variable aleatoria X distribuida según f_x . Si nos quedamos solamente con las varillas que miden más de 2cm, ¿cómo se distribuye la longitud de las varillas que quedan?"

Método para obtener la función de densidad de distribuciones condicionadas o "truncadas".

- 1) Hallar "a" (la probabilidad de la condición)
- 2) Crear una función exactamente igual a la original, pero restringiendo el dominio de modo de que no incluya los valores descartados.
- 3) Dividir por "a" todas las ramas de la nueva función.

Este método se aplica tanto a distribuciones discretas como continuas.

Recomendación práctica

Al igual que se dijo para la función de distribución acumulada, no conviene construir una función que no es necesaria, a menos que nos la pidan. Si lo que nos piden es una probabilidad, no necesitamos construir la función de densidad de la variable condicionada. En general todos los problemas que piden probabilidades de variables condicionadas se pueden resolver fácilmente por probabilidad condicional, sin necesidad alguna de la función de densidad condicionada. Veremos esto más adelante, plasmado en los ejemplos.

Resolución del ejemplo

$$f_X(x) = \begin{cases} 1/4 & 1 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

Si por ejemplo la distribución fuera

- Condición: $X > 2$

$$\int_{-\infty}^{+\infty} f_X(x) dx = \int_2^5 \frac{1}{4} dx = \frac{3}{4}$$

- $a = P(\text{condición}) = P(X > 2) = \frac{3}{4}$
- Función con el dominio restringido para que no incluya los valores descartados:

$$\begin{cases} 1/3 & 2 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

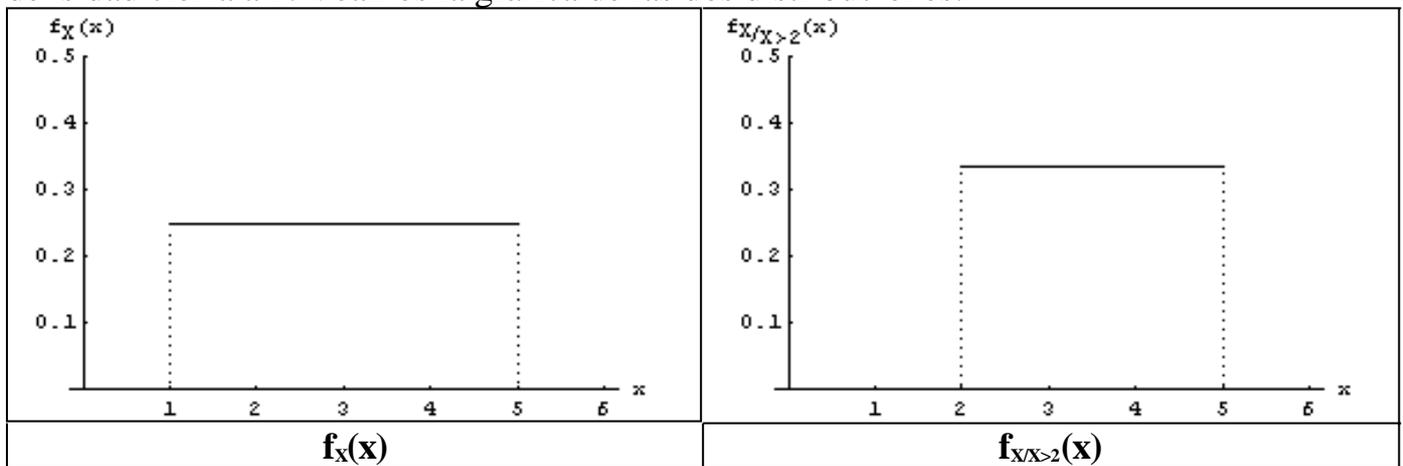
- Se dividen por "a" las ramas de la función:

$$\begin{cases} 1/4 & 2 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} 1/3 & 2 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

Lo que se obtuvo es la distribución de X, **condicionada a que $X > 2$** :

$$f_{X/X>2}(x) = \begin{cases} 1/3 & 2 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

Observamos que $X/X>2$ es efectivamente una distribución, ya que su función de densidad cierra a 1. Veamos la gráfica de las dos distribuciones:



Observamos que:

- 1) La nueva distribución tiene probabilidad nula para los valores que se descartaron.
- 2) La nueva distribución llega más alto que la anterior, porque el área debe cerrar a uno.

Ejemplo con una distribución discreta

Sea la variable aleatoria discreta X distribuida según:

$$P_X(x) = \begin{cases} 2/10 & x = 1 \\ 4/10 & x = 2 \\ 3/10 & x = 3 \\ 1/10 & x = 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

Un enunciado podría decir, por ejemplo: "Se tienen piezas de tipo 1, 2, 3 y 4, ubicadas, mezcladas, en una caja. El experimento consiste en tomar una pieza al azar de la caja. Hay un 20% de piezas tipo 1, 40% de tipo 2, 3% de tipo 3, y 10% de tipo 4. Luego alguien se toma el trabajo de quitar todas las piezas tipo 3 de la caja. ¿Cómo se distribuye X ahora?"

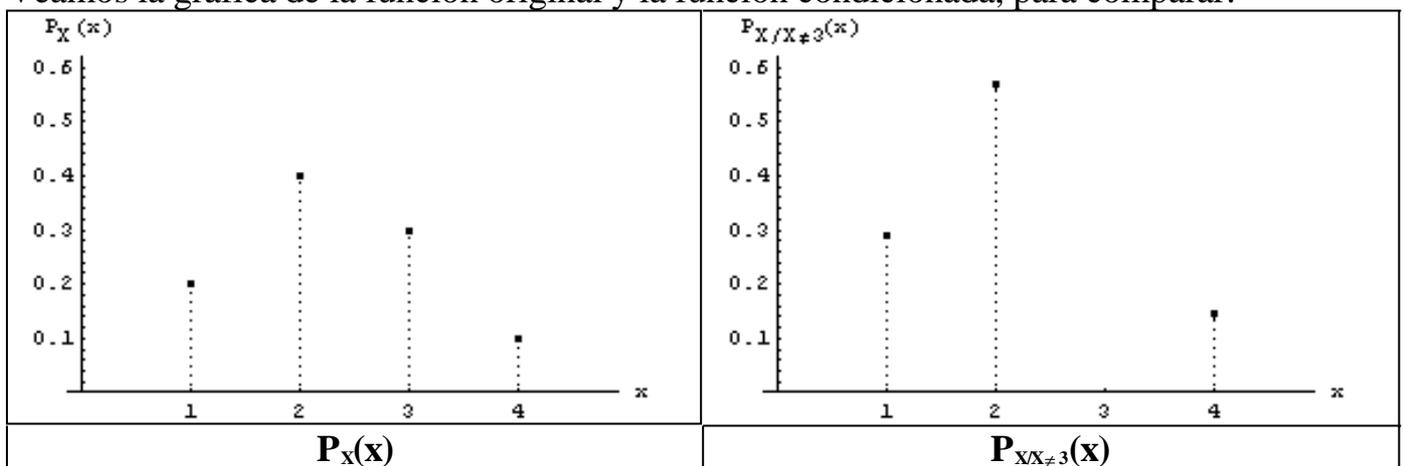
- Condición: $X = 3$
- $a = P(\text{condición}) = P(X \neq 3) = 7/10$
- Función con el dominio restringido para que no incluya los valores descartados:

$$P_X(x) = \begin{cases} 2/10 & x = 1 \\ 4/10 & x = 2 \\ 1/10 & x = 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

- Se dividen por 7/10 las ramas de la función. Lo que se obtiene es la distribución de X condicionada a que $X \neq 3$.

$$P_{X/X \neq 3}(x) = \begin{cases} 2/7 & x = 1 \\ 4/7 & x = 2 \\ 1/7 & x = 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

Veamos la gráfica de la función original y la función condicionada, para comparar:



Observamos que:

- 1) La distribución condicionada le asigna probabilidad cero al valor que fue descartado.
- 2) Los otros valores ahora tienen más probabilidad, para que el nuevo total también cierre a uno. Cuando se eliminó el 3, la probabilidad cerraba a 7/10. Por eso tuvimos que dividir todo por 7/10, para que cierre a uno.

3) Se mantuvo la proporción entre las probabilidades de los valores que quedaron. Es decir, $P(X=2)$ sigue siendo el doble de $P(X=1)$, y $P(X=1)$ sigue siendo el doble de $P(X=4)$. Está bien que esto sea así, porque las proporciones de piezas 1, 2 y 4 que quedaron no cambiaron. Por más que las piezas 3 ya no estén, sigue habiendo el doble de piezas 2 que de piezas 1, y sigue habiendo el doble de piezas 1 que de piezas 4.

Justificación del método

Pensemos en el ejemplo discreto que acabamos de ver. La probabilidad de sacar una pieza tipo 1 luego de haber eliminado las tipo 3, es:

$$P(X=1/X \neq 3)$$

Lo que estamos diciendo es que los tres renglones de $P_{X|X \neq 3}(x)$, es decir:

$P_{X|X \neq 3}(1)$, $P_{X|X \neq 3}(2)$ y $P_{X|X \neq 3}(4)$

En realidad son:

$$P(X=1/X \neq 3), P(X=2/X \neq 3) \text{ y } P(X=4/X \neq 3)$$

Veamos si efectivamente llegamos al mismo resultado. Tomemos $P_{X|X \neq 3}(1)$. Usando probabilidad condicional, queda:

$$P(X=1/X \neq 3) = \frac{P(X=1 \cap X \neq 3)}{P(X \neq 3)}$$

En el numerador, la condición $X=1$ "absorbe" a la condición $X \neq 3$, porque el suceso $X=1$ es un subconjunto del suceso $X \neq 3$. Luego la intersección es directamente el suceso $X=1$. Queda:

$$P(X=1/X \neq 3) = \frac{P(X=1)}{P(X \neq 3)}$$

Usando los datos de la distribución original, queda:

$$P(X=1/X \neq 3) = \frac{2/10}{7/10} = \frac{2}{7}$$

Vemos que llegamos al mismo resultado que con el método.

Para $X=2$, podemos hacer análogamente:

$$P(X=2/X \neq 3) = \frac{P(X=2 \cap X \neq 3)}{P(X \neq 3)} = \frac{P(X=2)}{P(X \neq 3)} = \frac{4/10}{7/10} = \frac{4}{7}$$

Y nuevamente obtuvimos el mismo resultado. Incluso para $X=3$ podemos hacer:

$$P(X=3/X \neq 3) = \frac{P(X=3 \cap X \neq 3)}{P(X \neq 3)} = \frac{P(\emptyset)}{P(X \neq 3)} = \frac{0}{7/10} = 0$$

Vemos que obtuvimos de manera formal lo que antes habíamos planteado intuitivamente: que la probabilidad de que X sea 3 en la nueva distribución debe ser cero. Cuando escribimos la nueva distribución, no usamos un renglón para indicar que la probabilidad del 3 es cero, porque directamente está contemplada en el \forall otro x .

Por último, para $X = 4$ obtenemos:

$$P\left(\frac{X = 4}{X \neq 3}\right) = \frac{P(X = 4 \cap X \neq 3)}{P(X \neq 3)} = \frac{P(X = 4)}{P(X \neq 3)} = \frac{1/10}{7/10} = \frac{1}{7}$$

Nuevamente obtuvimos lo esperado. Para generalizar lo que acabamos de hacer, podemos escribir:

$$P_{X/X \neq 3}(x) = P\left(\frac{X = x}{X \neq 3}\right) = \frac{P(X = x \cap X \neq 3)}{P(X \neq 3)}$$

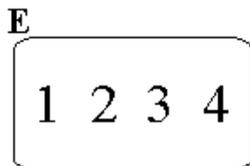
Y eso debe cumplirse para todos los valores posibles de x . Observemos que:

- 1) El numerador da la probabilidad original cuando $x \neq 3$, y cero cuando $x = 3$.
- 2) El denominador es una constante.
- 3) Vemos entonces que lo que estamos haciendo es tomar las probabilidades originales, ponerle cero a los valores descartados, y dividir por una constante.
- 4) Al dividir todo por la misma constante, estamos manteniendo las proporciones originales.
- 5) La constante que estamos usando para dividir es tal que la probabilidad de la nueva distribución también cierra a 1.

Vemos ahora que el método que enunciamos al principio de la sección es válido.

Otra forma de verlo

Otra forma de verlo es pensando en el espacio muestral. Nuestro espacio muestral es originalmente:



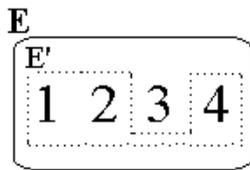
$$P_E(1) = 2/10$$

$$P_E(2) = 4/10$$

$$P_E(3) = 3/10$$

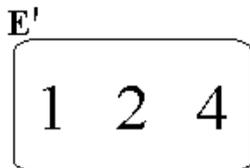
$$P_E(4) = 1/10$$

Ante el conocimiento de que el 3 no sale, tenemos que el conjunto de valores posibles es un espacio muestral más pequeño incluido en el original:



$$P_E(E') = 7/10$$

Luego, como se vio en el capítulo 1, para conocer las probabilidades de los valores que quedaron, referidos al nuevo espacio muestral, debemos dividir las originales por $P_E(E')$. El nuevo espacio muestral E' es:



$$P_{E'}(1) = \frac{P_E(1)}{P_E(E')} = \frac{2/10}{7/10} = \frac{2}{7}$$

$$P_{E'}(2) = \frac{P_E(2)}{P_E(E')} = \frac{4/10}{7/10} = \frac{4}{7}$$

$$P_{E'}(4) = \frac{P_E(4)}{P_E(E')} = \frac{1/10}{7/10} = \frac{1}{7}$$

Problemas típicos

1) Los huevos producidos en un determinado gallinero se clasifican, de acuerdo a sus tamaños, en 4 categorías: 1, 2, 3 y 4, que comprenden respectivamente el 5%, el 60%, el 20% y el 15% de los huevos. Para una partida de huevos, se pide:

a) Si se descartaran los huevos de calidad 4, por ser de calidad demasiado inferior, ¿cómo quedarían distribuidas las calidades de los huevos que no descartamos?

b) Si se descartan los huevos de calidad 4 como se indicó en a, ¿cuál es la probabilidad de elegir un huevo al azar y que sea de calidad 1?

Resuelva de dos formas distintas, y saque conclusiones.

- c) Si se venden los huevos de calidad 1 y 2, y no se venden los de calidad 3 y 4, ¿cómo se distribuyen los huevos vendidos? ¿cómo se distribuyen los huevos no vendidos?
- d) Si los huevos de 1^{ra} calidad se vendieron, y los huevos de 4^{ta} calidad se descartaron, ¿cómo se distribuyen los huevos que quedan?
- e) Si un cliente nos compra todos los huevos de calidad 2, ¿cómo se distribuyen los huevos que quedan?

Resolución

Comencemos por escribir la distribución original:

$$P_X(x) = \left\{ \begin{array}{ll} 0.05 & x=1 \\ 0.6 & x=2 \\ 0.2 & x=3 \\ 0.15 & x=4 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

a) Condición: huevo no descartado (es decir, calidad < 4)

$$P(\text{condición}) = P(X < 4) = P(X = 1) + P(X = 2) + P(X = 3) = 0.85 = a$$

$$P_{X/X < 4}(x) = \left\{ \begin{array}{ll} \frac{0.05}{a} & x=1 \\ \frac{0.6}{a} & x=2 \\ \frac{0.2}{a} & x=3 \\ \frac{a}{a} & \forall \text{ otro } x \end{array} \right\} = \left\{ \begin{array}{ll} 0.059 & x=1 \\ 0.706 & x=2 \\ 0.235 & x=3 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

b)

Forma 1: aprovechando la distribución condicionada que calculamos en el punto anterior:

$$P(X=1/X < 4) = P_{X/X < 4}(1) = 0,059$$

Forma 2: usando directamente probabilidad condicional:

$$P(X=1/X < 4) = \frac{P(X=1 \cap X < 4)}{P(X < 4)} = \frac{P(X=1 \cap (X=1 \cup X=2 \cup X=3))}{P(X=1 \cup X=2 \cup X=3)} =$$

$$= \frac{P(X=1)}{P(X=1 \cup X=2 \cup X=3)} = \frac{0.05}{0.85} = 0,059$$

Vemos que si lo que nos piden es una probabilidad, no hace falta encontrar la distribución condicionada. Podemos encontrar la respuesta usando simplemente probabilidad condicional.

c₁) Condición: huevo vendido

$$P(\text{condición}) = P(X = 1) + P(X = 2) = 0.65 = a$$

$$P_{X/X=1 \vee X=2}(x) = \left\{ \begin{array}{l} \frac{0.05}{a} \quad x=1 \\ \frac{0.6}{a} \quad x=2 \\ 0 \quad \forall \text{ otro } x \end{array} \right\} = \left\{ \begin{array}{l} 0.077 \quad x=1 \\ 0.923 \quad x=2 \\ 0 \quad \forall \text{ otro } x \end{array} \right\}$$

c₂) Condición: huevo no vendido

$$P(\text{condición}) = P(X = 3) + P(X = 4) = 0.35 = a$$

$$P_{X/X=3 \vee X=4}(x) = \left\{ \begin{array}{l} \frac{0.2}{a} \quad x=3 \\ \frac{0.15}{a} \quad x=4 \\ 0 \quad \forall \text{ otro } x \end{array} \right\} = \left\{ \begin{array}{l} 0.571 \quad x=3 \\ 0.429 \quad x=4 \\ 0 \quad \forall \text{ otro } x \end{array} \right\}$$

d) Condición: huevo ni vendido ni descartado

$$P(\text{condición}) = P(X = 2) + P(X = 3) = 0.8 = a$$

$$P_{X/X=2 \vee X=3}(x) = \left\{ \begin{array}{l} \frac{0.6}{a} \quad x=2 \\ \frac{0.2}{a} \quad x=3 \\ 0 \quad \forall \text{ otro } x \end{array} \right\} = \left\{ \begin{array}{l} 0.75 \quad x=2 \\ 0.25 \quad x=3 \\ 0 \quad \forall \text{ otro } x \end{array} \right\}$$

e) Condición: huevo de calidad que no sea 2

$$P(\text{condición}) = P(X \neq 2) = 1 - P(X = 2) = 0.4$$

$$P_{X/X \neq 2}(x) = \left\{ \begin{array}{l} \frac{0.05}{a} \quad x=1 \\ \frac{0.2}{a} \quad x=3 \\ \frac{0.15}{a} \quad x=4 \\ \frac{a}{0} \quad \forall \text{ otro } x \end{array} \right\} = \left\{ \begin{array}{l} 0.125 \quad x=1 \\ 0.5 \quad x=3 \\ 0.375 \quad x=4 \\ 0 \quad \forall \text{ otro } x \end{array} \right\}$$

2) La longitud en cm. de las varillas fabricadas por una máquina es una variable aleatoria X distribuida según:

$$f_X(x) = \begin{cases} \frac{1}{4} & 1 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

Se pide encontrar las distribución de la longitud de las varillas:

- que quedan, si se descartan las que miden menos de 2 cm.
- que quedan, si se descartan las que miden más de 4 cm.
- que quedan, si descartan las que miden menos de 2 cm. y las que miden más de 4 cm.
- que miden más de 4 cm.
- descartadas en el punto c.
- Calcular la probabilidad de que una varilla que mide más de 2 cm y menos de 4 cm, mida menos de 3.5 cm. Resolver de dos formas distintas y sacar conclusiones.

Resolución

a) Condición: $X > 2$

$$a = P(\text{condición}) = P(X > 2) = \int_2^{+\infty} f_X(x) dx = \int_2^5 \frac{1}{4} dx = \frac{3}{4}$$

$$f_{X/X>2}(x) = \begin{cases} \frac{1/4}{a} & 2 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} \frac{1}{3} & 2 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

b) Condición: $X < 4$

$$a = P(X < 4) = \int_{-\infty}^4 f_X(x) dx = \int_1^4 \frac{1}{4} dx = \frac{3}{4}$$

$$f_{X/X<4}(x) = \begin{cases} \frac{1/4}{a} & 1 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} \frac{1}{3} & 1 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

c) Condición: varillas que quedan = $X > 2 \wedge X < 4$

$$a = P(X > 2 \wedge X < 4) = \int_2^4 f_X(x) dx = \int_2^4 \frac{1}{4} dx = \frac{1}{2}$$

$$f_{X/2<X<4}(x) = \begin{cases} \frac{1/4}{a} & 2 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} \frac{1}{2} & 2 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

d) Condición: $X > 4$

$$a = P(X > 4) = \int_4^{+\infty} f_X(x) dx = \int_4^5 \frac{1}{4} dx = \frac{1}{4}$$

$$f_{X/X>4}(x) = \begin{cases} \frac{1}{4} & 4 < x < 5 \\ \frac{a}{0} & \forall \text{ otro } x \end{cases} = \begin{cases} 1 & 4 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

e) Condición: descartadas en el punto c = las que son mayores de 4 o menores de 2

$$a = P(X < 2 \cup X > 4) = \int_{-\infty}^2 f_X(x) dx + \int_4^{+\infty} f_X(x) dx = \int_1^2 \frac{1}{4} dx + \int_4^5 \frac{1}{4} dx = \frac{1}{2}$$

$$f_{X/X<2 \vee X>4}(x) = \begin{cases} \frac{1/4}{a} & 1 < x < 2 \vee 4 < X < 5 \\ \frac{0}{0} & \forall \text{ otro } x \end{cases} = \begin{cases} \frac{1}{2} & 1 < x < 2 \vee 4 < X < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

f)

Forma 1:

$$P\left(\frac{X < 3.5}{2} < X < 4\right) = \int_{-\infty}^{\frac{3.5}{2}} f_{X/X < 2 \vee X > 4}(x) dx = \int_2^{\frac{3.5}{2}} \frac{1}{2} dx = 0.75$$

Forma 2:

$$P\left(\frac{X < 3.5}{2} < X < 4\right) = \frac{P(X < 3.5 \cap 2 < X < 4)}{P(2 < X < 4)} = \frac{P(2 < X < 3.5)}{P(2 < X < 4)} = \frac{\int_2^{\frac{3.5}{2}} \frac{1}{4} dx}{\int_2^4 \frac{1}{4} dx} = \frac{3/8}{4/8} = 0.75$$

Vemos que si lo que nos piden es una probabilidad, no hace falta encontrar la distribución condicionada. Podemos encontrar la respuesta usando simplemente la definición de probabilidad condicional (y de esa forma lo que estamos obteniendo es solamente el renglón que necesitamos de la distribución condicionada).

3) Idem 2, con las varillas distribuidas según:

$$f_X(x) = \begin{cases} \frac{(x-1)^2}{4} & 1 \leq x \leq 3 \\ \frac{x-3}{6} & 3 < x \leq 5 \\ \frac{0}{0} & \forall \text{ otro } x \end{cases}$$

Resolución

a) Condición: $X > 2$

$$a = P(\text{condición}) = P(X > 2) = \int_2^3 f_X(x) dx + \int_3^5 \frac{x-3}{6} dx = 0.917$$

$$f_{X/X>2}(x) = \left\{ \begin{array}{ll} \frac{1}{a} \frac{(x-1)^2}{4} & 2 < x \leq 3 \\ \frac{1}{a} \frac{x-3}{6} & 3 < x \leq 5 \\ 0 & \forall \text{ otro } x \end{array} \right\} = \left\{ \begin{array}{ll} \frac{(x-1)^2}{3.67} & 2 < x \leq 3 \\ \frac{x-3}{5.5} & 3 < x \leq 5 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

b) Condición: $X < 4$

$$a = P(X < 4) = \int_{-\infty}^4 f_X(x) dx = \int_1^3 \frac{(x-1)^2}{4} dx + \int_3^4 \frac{x-3}{6} dx = 0.75 = \frac{3}{4}$$

$$f_{X/X<4}(x) = \left\{ \begin{array}{ll} \frac{1}{a} \frac{(x-1)^2}{4} & 1 \leq x \leq 3 \\ \frac{1}{a} \frac{x-3}{6} & 3 < x < 4 \\ 0 & \forall \text{ otro } x \end{array} \right\} = \left\{ \begin{array}{ll} \frac{(x-1)^2}{3} & 1 \leq x \leq 3 \\ \frac{x-3}{4.5} & 3 < x < 4 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

c) Condición: varillas que quedan = $X > 2 \wedge X < 4$

$$a = P(X > 2 \wedge X < 4) = \int_2^4 f_X(x) dx = \int_2^3 \frac{(x-1)^2}{4} dx + \int_3^4 \frac{x-3}{6} dx = \frac{2}{3}$$

$$f_{X/2<X<4}(x) = \left\{ \begin{array}{ll} \frac{1}{a} \frac{(x-1)^2}{4} & 2 < x \leq 3 \\ \frac{1}{a} \frac{x-3}{6} & 3 < x < 4 \\ 0 & \forall \text{ otro } x \end{array} \right\} = \left\{ \begin{array}{ll} \frac{3}{8} (x-1)^2 & 2 < x \leq 3 \\ \frac{1}{4} (x-3) & 3 < x < 4 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

d) Condición: $X > 4$

$$a = P(X > 4) = \int_4^{\infty} f_X(x) dx = \int_4^5 \frac{x-3}{6} dx = \frac{1}{4}$$

$$f_{X/X>4}(x) = \left\{ \begin{array}{ll} \frac{1}{a} \frac{x-3}{6} & 4 < X \leq 5 \\ 0 & \forall \text{ otro } x \end{array} \right\} = \left\{ \begin{array}{ll} \frac{2}{3} (x-3) & 4 < X \leq 5 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

e) Condición: descartadas en el punto c = las que son mayores de 4 o menores de 2

$$a = P(X < 2 \cup X > 4) = \int_{-\infty}^2 f_X(x) dx + \int_4^{\infty} f_X(x) dx = \int_1^2 \frac{(x-1)^2}{4} dx + \int_4^5 \frac{x-3}{6} dx = \frac{1}{3}$$

$$f_{X/2 < X < 4} (x) = \left\{ \begin{array}{ll} \frac{1}{a} \frac{(x-1)^2}{4} & 1 \leq x < 2 \\ \frac{1}{a} \frac{x-3}{6} & 4 < x \leq 5 \\ 0 & \forall \text{ otro } x \end{array} \right\} = \left\{ \begin{array}{ll} \frac{3}{4} (x-1)^2 & 1 \leq x < 2 \\ \frac{1}{2} (x-3) & 4 < x \leq 5 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

f)

Forma 1:

$$P\left(\frac{X < 3.5}{2 < X < 4}\right) = \int_{-\infty}^{3.5} f_{X/2 < X < 4} (x) dx = \int_2^{3.5} \frac{1}{2} dx = 0.75$$

Forma 2:

$$P\left(\frac{X < 3.5}{2 < X < 4}\right) = \frac{P(X < 3.5 \cap 2 < X < 4)}{P(2 < X < 4)} = \frac{P(2 < X < 3.5)}{P(2 < X < 4)} = \frac{\int_2^3 \frac{(x-1)^2}{4} dx + \int_3^{3.5} \frac{x-3}{6} dx}{\int_2^3 \frac{(x-1)^2}{4} dx + \int_3^4 \frac{x-3}{6} dx} = \frac{0.6042}{0.6667} = 0.9063$$

Vemos que si lo que nos piden es una probabilidad, no hace falta encontrar la distribución condicionada. Podemos encontrar la respuesta usando simplemente probabilidad condicional.

Es frecuente encontrar ejemplos de variables aleatorias condicionadas combinadas con variables aleatorias mezcla. Ejemplos de tales casos son abordados en la siguiente sección: "Variable aleatoria mezcla"



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/process>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 4 de mayo de 2004_

Variable aleatoria mezcla

"Las máquinas A, B y C producen piezas cuyas longitudes están distribuidas respectivamente según f_{XA} , f_{XB} y f_{XC} . Las máquinas producen respectivamente el 20%, el 50% y el 30% del total de unidades producidas. ¿Cómo se distribuyen las longitudes de las piezas producidas?"

Método para obtener la función de densidad de la variable aleatoria mezcla

1) Averiguar las probabilidades de cada uno de los orígenes:

$$P(A_1), P(A_2), \dots, P(A_n)$$

2) Averiguar la distribución de los elementos provenientes de cada uno de los orígenes:

$$f_{X1}, f_{X2}, \dots, f_{Xn}.$$

3) Hacer una lista de todos los puntos que dividen las ramas de las funciones de densidad de todos los orígenes.

4) Para cada intervalo:

$$f_{XMEZCLA} = P(A_1) f_{X1}(x) + P(A_2) f_{X2}(x) + \dots + P(A_n) f_{Xn}(x)$$

5) Armar la $f_{XMEZCLA}$.

Recomendaciones prácticas

- Al igual que se dijo para la función de distribución acumulada y para la distribución de las variables aleatorias condicionadas, no conviene construir una función que no es necesaria, a menos que nos la pidan. Si lo que nos piden es una probabilidad, no necesitamos construir la función de densidad de la variable

aleatoria mezcla. En general todos los problemas que piden probabilidades de variables aleatorias mezcla se pueden resolver fácilmente por probabilidad condicional, sin necesidad alguna de la función de densidad de la mezcla. Veremos esto más adelante, plasmado en los ejemplos.

- Aunque el método no lo requiere, es conveniente, para tener menos probabilidades de equivocarse, hacer un diagrama del estilo del que se hará en la resolución del ejemplo.

Resolución del ejemplo

Si por ejemplo las distribuciones fueran:

$$f_{x_A}(x) = \begin{cases} \frac{1}{4} & 2 < x < 6 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_{x_B}(x) = \begin{cases} \frac{1}{3} & 4 < x < 7 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_{x_C}(x) = \begin{cases} \frac{1}{4} & 1 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

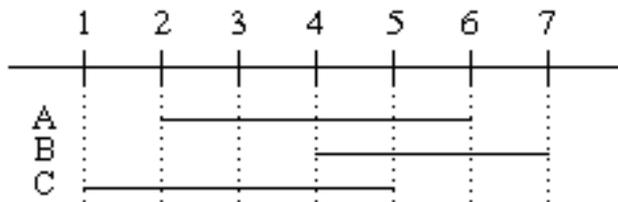
y era dato que:

$$P(A) = 0,2 \quad P(B) = 0,5 \quad P(C) = 0,3$$

Los puntos que separan las ramas de las 3 funciones de densidad son: 2, 6, 4, 7, 1, 5.

Ordenados quedan: 1, 2, 4, 5, 6, 7.

Hagamos un esquema para darnos cuenta de en qué intervalos "aportan" cada uno de los orígenes:



En cada intervalo aplicaremos:

$$f_{XMEZCLA} = P(A) f_{x_A}(x) + P(B) f_{x_B}(x) + P(C) f_{x_C}(x)$$

- Intervalo $-\infty < x < 1$

$$f_{XMEZCLA} = 0,2 \cdot 0 + 0,5 \cdot 0 + 0,3 \cdot 0 = 0$$

Es lógico que de 0, porque no hay aportes.

- Intervalo $1 < x < 2$

$$f_{XMEZCLA} = 0,2 \cdot 0 + 0,5 \cdot 0 + 0,3 \cdot \frac{1}{4} = \frac{3}{40}$$

- Intervalo $2 < x < 4$

$$f_{XMEZCLA} = 0,2 \cdot \frac{1}{4} + 0,5 \cdot 0 + 0,3 \cdot \frac{1}{4} = \frac{1}{8}$$

- Intervalo $4 < x < 5$

$$f_{XMEZCLA} = 0,2 \cdot \frac{1}{4} + 0,5 \cdot \frac{1}{3} + 0,3 \cdot \frac{1}{4} = \frac{7}{24}$$

- Intervalo $5 < x < 6$

$$f_{XMEZCLA} = 0,2 \cdot \frac{1}{4} + 0,5 \cdot \frac{1}{3} + 0,3 \cdot 0 = \frac{13}{60}$$

- Intervalo $6 < x < 7$

$$f_{XMEZCLA} = 0,2 \cdot 0 + 0,5 \cdot \frac{1}{3} + 0,3 \cdot 0 = \frac{1}{6}$$

- Intervalo $7 < x < +\infty$

$$f_{XMEZCLA} = 0,2 \cdot 0 + 0,5 \cdot 0 + 0,3 \cdot 0 = 0$$

Ahora armamos la $f_{XMEZCLA}$:

$$f_{X_{\text{MEZCLA}}}(x) = \begin{cases} 3/40 & 1 < x < 2 \\ 1/8 & 2 < x < 4 \\ 7/24 & 4 < x < 5 \\ 13/60 & 5 < x < 6 \\ 1/6 & 6 < x < 7 \\ 0 & \forall \text{ otro } x \end{cases}$$

Justificación del método

Comenzaremos por buscar la $F_{X_{\text{MEZCLA}}}(x)$:

Por la definición de función de probabilidad acumulada, podemos escribir:

$$F_{X_m}(x) = P(X_m \leq x)$$

Luego tomaremos en vez del suceso $X_m \leq x$ su intersección con el espacio muestral, lo cual nos dará un suceso equivalente, por lo cual no se modificará la probabilidad:

$$P(X_m \leq x) = P(X_m \leq x \cap E)$$

Escribiremos el espacio muestral como los sucesos: "que un determinado elemento venga del origen 1 ó que venga del origen 2 ó ... ó que venga del origen n":

$$P(X_m \leq x \cap E) = P(X_m \leq x \cap (A_1 \cup \dots \cup A_n))$$

Luego distribuimos la intersección respecto de la unión, y queda:

$$P(X_m \leq x \cap (A_1 \cup \dots \cup A_n)) = P((X_m \leq x \cap A_1) \cup \dots \cup (X_m \leq x \cap A_n))$$

Como un elemento no puede venir de dos orígenes, entonces los sucesos son disjuntos, y podemos reemplazar la probabilidad de la unión por la suma de las probabilidades:

$$P((X_m \leq x \cap A_1) \cup \dots \cup (X_m \leq x \cap A_n)) = P(X_m \leq x \cap A_1) + \dots + P(X_m \leq x \cap A_n)$$

Podemos escribir las probabilidades de intersecciones usando probabilidades condicionales:

$$P(X_m \leq x \cap A_1) + \dots + P(X_m \leq x \cap A_n) = P(A_1)P(X_m \leq x / A_1) + \dots + P(A_n)P(X_m \leq x / A_n)$$

La probabilidad de que un determinado elemento sea menor a un determinado valor, sabiendo que vino de un determinado origen, es la función de probabilidad acumulada de ese origen, evaluada en el valor, con lo cual:

$$P(A_1)P(X_m \leq x / A_1) + \dots + P(A_n)P(X_m \leq x / A_n) = P(A_1)F_{X_1}(x) + \dots + P(A_n)F_{X_n}(x)$$

Entonces obtuvimos que:

$$F_{X_m}(x) = P(A_1)F_{X_1}(x) + \dots + P(A_n)F_{X_n}(x)$$

Y como:

$$f_{X_m}(x) = \frac{dF_{X_m}(x)}{dx}$$

Entonces:

$$f_{X_m}(x) = P(A_1)f_{X_1}(x) + \dots + P(A_n)f_{X_n}(x)$$

Lo cual vale $\forall x$. Entonces hacemos la lista de los puntos que dividan todas las

ramas de cada una de las f_{x_i} , para que si trabajamos en cada uno de los intervalos que quedan determinados entre esos puntos, no cambie la definición de ninguna de las funciones de densidad.

Problemas típicos

1) Los pesos de los duraznos, ciruelas y naranjas se distribuyen respectivamente en decagramos según:

$$f_{x_N}(x) = \begin{cases} x-5 & 5 < x < 6 \\ 7-x & 6 < x < 7 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_{x_C}(x) = \begin{cases} x-4 & 4 < x < 5 \\ 6-x & 5 < x < 6 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_{x_D}(x) = \begin{cases} \frac{1}{3} & 4 < x < 7 \\ 0 & \forall \text{ otro } x \end{cases}$$

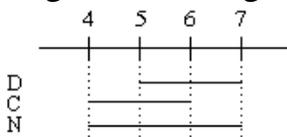
Si se mezclan 20% de naranjas, 30% de ciruelas y 50% de duraznos.

- ¿Cómo se distribuye el peso de una fruta elegida al azar?**
- ¿Cuál es la probabilidad de que una fruta elegida al azar pese menos de 6 decagramos? Resuelva de dos formas distintas y extraiga conclusiones.**
- Si se extrae una fruta al azar y pesa menos de 6 decagramos, ¿cuál es la probabilidad de que sea un durazno? Extraiga conclusiones.**

Resolución

a) Puntos que dividen ramas: 4, 5, 6, 7

Hagamos un diagrama que nos permita ir verificando lo que escribimos:



En cada intervalo aplicaremos:

$$f_{X_{MEZCLA}} = P(A) f_{x_D}(x) + P(B) f_{x_C}(x) + P(C) f_{x_N}(x)$$

• Intervalo $-\infty < x < 4$

$$f_{X_{MEZCLA}} = 0,2 \cdot 0 + 0,3 \cdot 0 + 0,5 \cdot 0 = 0$$

• Intervalo $4 < x < 5$

$$f_{X_{MEZCLA}} = 0,2 \cdot 0 + 0,3 \cdot (x-4) + 0,5 \cdot \frac{1}{3} = \frac{3}{10}(x-4) + \frac{1}{6}$$

• Intervalo $5 < x < 6$

$$f_{X_{MEZCLA}} = 0,2 \cdot (x-5) + 0,3 \cdot (6-x) + 0,5 \cdot \frac{1}{3} = \frac{1}{5}(x-5) + \frac{3}{10}(6-x) + \frac{1}{6}$$

• Intervalo $6 < x < 7$

$$f_{X_{MEZCLA}} = 0,2 \cdot (7-x) + 0,3 \cdot 0 + 0,5 \cdot \frac{1}{3} = \frac{1}{5}(7-x) + \frac{1}{6}$$

• Intervalo $7 < x < +\infty$

$$f_{X_{MEZCLA}} = 0,2 \cdot 0 + 0,3 \cdot 0 + 0,5 \cdot 0 = 0$$

Ahora armamos la $f_{X_{MEZCLA}}$:

$$f_{X_{\text{MEZCLA}}}(x) = \begin{cases} \frac{3}{10}(x-4) + \frac{1}{6} & 4 < x < 5 \\ \frac{1}{5}(x-5) + \frac{3}{10}(6-x) + \frac{1}{6} & 5 < x < 6 \\ \frac{1}{5}(7-x) + \frac{1}{6} & 6 < x < 7 \\ 0 & \forall \text{ otro } x \end{cases}$$

b) Primera forma:

$$P(X_m < 6) = \int_{-\infty}^{\infty} f_{X_m} dx = \int_4^5 \left(\frac{3}{10}(x-4) + \frac{1}{6} \right) dx + \int_5^6 \left(\frac{1}{5}(x-5) + \frac{3}{10}(6-x) + \frac{1}{6} \right) dx = \frac{11}{15}$$

Segunda forma:

Usamos la misma técnica de intersección con el espacio muestral que usamos para demostrar la fórmula:

$$\begin{aligned} P(X_m < 6) &= P(X_m < 6 \cap E) = P(X_m < 6 \cap (N \cup C \cup D)) = P((X_m < 6 \cap N) \\ &\cup (X_m < 6 \cap C) \cup (X_m < 6 \cap D)) = P(X_m < 6 \cap N) + P(X_m < 6 \cap C) + P(X_m \\ &< 6 \cap D) = P(N) P(X_m < 6 / N) + P(C) P(X_m < 6 / C) + P(D) P(X_m < 6 / D) = \\ &= 0,2 \int_{-\infty}^{\infty} f_{X_N} dx + 0,3 \int_{-\infty}^{\infty} f_{X_C} dx + 0,5 \int_{-\infty}^{\infty} f_{X_D} dx = \frac{11}{15} \end{aligned}$$

En la segunda forma no necesitamos la distribución de la mezcla. Si no nos hubieran pedido la función de densidad, no habría valido la pena hacerla.

$$c) P\left(\frac{D}{X_m} < 6\right) = \frac{P\left(\frac{X_m < 6}{D}\right)P(D)}{P(X_m < 6)}$$

El denominador ya lo calculamos en el punto anterior.

P(D) es dato y vale 0,5.

$$P\left(\frac{X_m < 6}{D}\right) = \int_{-\infty}^{\infty} f_{X_D} dx = \int_4^6 \frac{1}{3} dx = \frac{2}{3}$$

$$P\left(\frac{D}{X_m} < 6\right) = \frac{\frac{2}{3} \cdot \frac{1}{2}}{\frac{11}{15}} = \frac{5}{11}$$

Nuevamente la conclusión es que si lo que se pide es calcular probabilidades, no hace falta encontrar la función de densidad de la variable aleatoria mezcla.

2) En un parque de diversiones, la altura de las personas que quieren subirse a determinada atracción mecánica es una variable aleatoria X distribuida en metros según:

$$f_{X_N}(x) = \begin{cases} \frac{25}{4}(x-1) & 1 < x < 1.2 \\ \frac{5}{4} & 1.2 < x < 1.8 \\ \frac{25}{4}(2-x) & 1.8 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

El empleado del parque no puede dejar subir a nadie que mida menos de 1.40m. Sin embargo, el 10% de los que miden menos de 1.40m igualmente logran escabullirse.

¿Cómo se distribuyen las alturas de las personas que se suben al juego?

Resolución

Este problema tiene dos dificultades adicionales:

- la mezcla es entre variables que a su vez proceden de distribuciones condicionadas.
- el cálculo de las probabilidades de los orígenes no es inmediato.

Comencemos por ver que va a haber que hacer una mezcla entre dos variables aleatorias: la de las personas de más de 1.40m y que entraron legítimamente, y las de las personas de menos de 1.40m, que se escabulleron:

$$f_{X_{MEZCLA}} = P(\text{entró escabullido}) f_{X_{bajos}}(x) + P(\text{entró legítimamente}) f_{X_{altos}}(x)$$

Entonces necesitaremos las distribuciones condicionadas:

- Personas de menos de 1.40m:

$$P(X < 1.4) = \int_1^{1.2} \frac{25}{4}(x-1)dx + \int_{1.2}^{1.4} \frac{5}{4} dx = \frac{3}{8}$$

$$\Rightarrow f_{X/X < 1.4}(x) = \begin{cases} \frac{8}{3} \frac{25}{4}(x-1) & 1 < x < 1.2 \\ \frac{8}{3} \frac{5}{4} & 1.2 < x < 1.4 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} \frac{50}{3}(x-1) & 1 < x < 1.2 \\ \frac{10}{3} & 1.2 < x < 1.4 \\ 0 & \forall \text{ otro } x \end{cases} = f_{X_{bajos}}$$

- Personas de más de 1.40m:

$$P(X > 1.4) = 1 - \frac{3}{8} = \frac{5}{8}$$

$$\Rightarrow f_{X/X > 1.4}(x) = \begin{cases} \frac{8}{5} \frac{5}{4} & 1.4 < x < 1.8 \\ \frac{8}{5} \frac{25}{4}(2-x) & 1.8 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} 2 & 1.4 < x < 1.8 \\ 10(2-x) & 1.8 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases} = f_{X_{altos}}$$

Y ahora necesitamos calcular las probabilidades. Mostraremos 3 formas de hacerlo: con proporciones, con probabilidad condicional, y con un diagrama del espacio muestral:

- Con proporciones, se hace así:

Por ejemplo, de cada 80 personas que se presentan, 50 son altas y 30 son bajas (según las probabilidades que acabamos de calcular). Las 50 que se presentan y son altas, entran directamente. De las 30 que son bajas, el 10% entra, con lo cual 3 entran. Entonces de las 80 que se presentaron, entran 53, de las cuales 50 son altas y 3 son bajas.

Luego la probabilidad de que una persona que entró lo haya hecho legítimamente es $50/53$, y la probabilidad de que una persona que entró lo haya hecho escabulléndose es $3/53$.

- Aplicando probabilidad condicional, se hace así:

$$P(\text{bajo} / \text{entró}) = \frac{P(\text{entró} / \text{bajo}) P(\text{bajo})}{P(\text{entró})} = \frac{10\% \frac{3}{8}}{P(\text{entró})} = \frac{3}{80 P(\text{entró})}$$

$$P(\text{alto} / \text{entró}) = \frac{P(\text{entró} / \text{alto}) P(\text{alto})}{P(\text{entró})} = \frac{100\% \frac{5}{8}}{P(\text{entró})} = \frac{5}{8 P(\text{entró})}$$

Y como todas las personas que entran son necesariamente altas o bajas, entonces:

$$P(\text{baja} / \text{entró}) + P(\text{alta} / \text{entró}) = 1$$

Con lo cual:

$$\frac{3}{80 P(\text{entró})} + \frac{5}{8 P(\text{entró})} = 1 \Rightarrow \frac{53}{80 P(\text{entró})} = 1 \Rightarrow P(\text{entró}) = \frac{53}{80}$$

Luego:

$$P(\text{baja} / \text{entró}) = \frac{3}{80 P(\text{entró})} = \frac{3 \cdot 80}{80 \cdot 53} = \frac{3}{53}$$

$$P(\text{alta} / \text{entró}) = \frac{5}{8 P(\text{entró})} = \frac{5 \cdot 80}{8 \cdot 53} = \frac{50}{53}$$

- Haciendo un diagrama del espacio muestral, se hace así:

La parte con líneas gruesas en el centro es el espacio muestral, dividido en las 4 particiones posibles: los que entran y son altos, los que entran y son bajos, los que no entran y son altos, y los que no entran y son bajos.

	entran	no entran	
altos			
bajos			

Completaremos los casilleros con las siguientes probabilidades, según las vayamos deduciendo:

	entran	no entran	
altos	$P(\text{alto} \wedge \text{entra})$	$P(\text{alto} \wedge \text{noentra})$	$P(\text{alto})$
bajos	$P(\text{bajo} \wedge \text{entra})$	$P(\text{bajo} \wedge \text{noentra})$	$P(\text{bajo})$
	$P(\text{entra})$	$P(\text{noentra})$	

Comenzamos por colocar las probabilidades que ya teníamos calculadas de cuando buscamos las distribuciones condicionadas:

	entran	no entran	
altos	$P(\text{alto} \wedge \text{entra})$	$P(\text{alto} \wedge \text{noentra})$	$5/8$
bajos	$P(\text{bajo} \wedge \text{entra})$	$P(\text{bajo} \wedge \text{noentra})$	$3/8$
	$P(\text{entra})$	$P(\text{noentra})$	

Y sabemos que todos los altos que se presentan entran, con lo cual $P(\text{alto} \wedge \text{noentra}) = 0$, por lo tanto:

	entran	no entran	
altos	$P(\text{alto} \wedge \text{entra})$	0	$5/8$
bajos	$P(\text{bajo} \wedge \text{entra})$	$P(\text{bajo} \wedge \text{noentra})$	$3/8$
	$P(\text{entra})$	$P(\text{noentra})$	

Y luego como podemos ver en el diagrama, $P(\text{alto} \wedge \text{entra}) + P(\text{alto} \wedge \text{noentra}) = 5/8$, con lo cual:

	entran	no entran	
altos	$5/8$	0	$5/8$
bajos	$P(\text{bajo} \wedge \text{entra})$	$P(\text{bajo} \wedge \text{noentra})$	$3/8$
	$P(\text{entra})$	$P(\text{noentra})$	

Sabemos que de los bajos, 10% entra y 90% no entra. El 10% de $3/8$ es $3/80$, y el 90% es $27/80$. Entonces:

	entran	no entran	
altos	$5/8$	0	$5/8$
bajos	$3/80$	$27/80$	$3/8$
	$P(\text{entra})$	$P(\text{noentra})$	

Y ahora sumamos para terminar de completar la tabla:

	entran	no entran	
altos	$5/8$	0	$5/8$
bajos	$3/80$	$27/80$	$3/8$
	$53/80$	$27/80$	

Ahora calculamos

$$P\left(\frac{\text{bajo}}{\text{entró}}\right) = \frac{P(\text{bajo} \wedge \text{entró})}{P(\text{entró})} = \frac{3/80}{53/80} = \frac{3}{53}$$

$$P\left(\frac{\text{alto}}{\text{entró}}\right) = \frac{P(\text{alto} \wedge \text{entró})}{P(\text{entró})} = \frac{5/8}{53/80} = \frac{50}{53}$$

Luego de conseguidas por cualquiera de los tres métodos las probabilidades, buscamos los puntos que dividen las ramas de las funciones de densidad:

1, 1.2, 1.4, 1.8, 2

Ahora trabajamos intervalo por intervalo, usando:

$$f_{\text{XMEZCLA}} = P(\text{entró escabullido}) f_{\text{Xbajos}}(x) + P(\text{entró legítimamente}) f_{\text{Xaltos}}(x)$$

• $-\infty < x < 1$

$$f_{\text{XMEZCLA}} = \frac{3}{53} 0 + \frac{50}{53} 0 = 0$$

- $1 < x < 1.2$

$$f_{\text{XMEZCLA}} = \frac{3}{53} \frac{50}{3} (x - 1) + \frac{50}{53} 0 = \frac{50}{53} (x - 1)$$

- $1.2 < x < 1.4$

$$f_{\text{XMEZCLA}} = \frac{3}{53} \frac{10}{3} + \frac{50}{53} 0 = \frac{10}{53}$$

- $1.4 < x < 1.8$

$$f_{\text{XMEZCLA}} = \frac{3}{53} 0 + \frac{50}{53} 2 = \frac{100}{53}$$

- $1.8 < x < 2$

$$f_{\text{XMEZCLA}} = \frac{3}{53} 0 + \frac{50}{53} 10(2 - x) = \frac{500}{53} (2 - x)$$

- $2 < x < +\infty$

$$f_{\text{XMEZCLA}} = \frac{3}{53} 0 + \frac{50}{53} 0 = 0$$

Finalmente, armamos la función de densidad:

$$f_{\text{XMEZCLA}} = \left\{ \begin{array}{ll} \frac{50}{53} (x - 1) & 1 < x < 1.2 \\ \frac{10}{53} & 1.2 < x < 1.4 \\ \frac{100}{53} & 1.4 < x < 1.8 \\ \frac{500}{53} (2 - x) & 1.8 < x < 2 \\ 0 & \forall \text{ otro } x \end{array} \right.$$

4) Un artesano recolecta varillas cuya longitud en cm. es la variable aleatoria X distribuida según:

$$f_x(x) = \left\{ \begin{array}{ll} \frac{2}{25} (x - 1) & 1 < x < 6 \\ 0 & \forall \text{ otro } x \end{array} \right.$$

Como las varillas que miden más de 4 cm no le sirven, por ser demasiado grandes, las corta por la mitad.

a) ¿Cómo se distribuyen las varillas que le quedan?

b) Si de las varillas que le quedan selecciona las que miden $2 \pm 0,1$ cm., ¿cómo se distribuyen las que selecciona?

Resolución:

a) Este ejercicio es lo más complicado que estudiaremos en cuanto a variable aleatoria mezcla, ya que además comprende:

- distribuciones condicionadas
- probabilidades de orígenes
- cambio de variables

Comencemos por ver que la mezcla será entre las varillas que miden menos de 4 cm, y las que resultaron de cortar por la mitad varillas que medían originalmente más de 4 cm.

La distribución de las varillas que miden menos de 4 cm se puede obtener fácilmente truncando la distribución original condicionándola a $X < 4$. Las otras, resultarán del cambio de variables de cortar por la mitad varillas que provienen de otra distribución condicionada.

Comencemos por obtener las distribuciones condicionadas:

$$P(X < 4) = 9/25$$

$$f_{x/x < 4}(x) = \begin{cases} \frac{25}{9} \frac{2}{25} (x-1) & 1 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} \frac{2}{9} (x-1) & 1 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

=>

$$P(X > 4) = 16/25$$

$$f_{x/x > 4}(x) = \begin{cases} \frac{25}{16} \frac{2}{25} (x-1) & 4 < x < 6 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} \frac{2}{16} (x-1) & 4 < x < 6 \\ 0 & \forall \text{ otro } x \end{cases}$$

=>

Por simplicidad, a la longitud de las varillas de más de 4 cm. la llamaremos Y, y Z será la longitud de tales varillas cortadas por la mitad.

$$Z = Y / 2$$

Hacemos el cambio de variables:

La derivada de la transformación es 1/2. La transformación inversa es $Y = 2Z$.

Los puntos que dividen las ramas de f_Y (es decir, de $f_{x/x > 4}$) son el 4 y el 6.

No hay puntos que dividan ramas de la derivada, y tampoco puntos en los que la derivada cambie de signo.

Entonces el único intervalo a estudiar es $4 < Y < 6$



$$Y = 4 \Rightarrow Z = 2$$

$$Y = 6 \Rightarrow Z = 3$$

con lo cual dicho intervalo aportará sobre el intervalo $2 < Z < 3$

$$f_Z(z) = \frac{f_Y(y)}{\left| \frac{dz}{dy} \right|} = \frac{\frac{2}{16} (y-1)}{\frac{1}{2}} = \frac{1}{4} (y-1) = \frac{1}{4} (2z-1)$$

La distribución de z queda:

$$f_z(z) = \begin{cases} \frac{1}{4}(2z-1) & 2 < z < 3 \\ 0 & \forall \text{ otro } z \end{cases}$$

Ahora calcularemos las probabilidades de los orígenes. Usaremos el método de pensar en las proporciones, aunque también podría usarse cualquiera de los otros 2. De cada 25 varillas, 9 quedarán como están, y 16 se cortarán. Pero las 16 que se cortan se transforman en 32. Entonces por cada 25 varillas, quedarán $9+32 = 41$ varillas. De cada 41 varillas de las que quedan, 9 eran cortas originalmente, y 32 proceden de las largas cortadas por la mitad.

Las que eran cortas originalmente son las que están distribuidas según $f_{X|X<4}$. Las que resultaron de cortar las largas están distribuidas según la f_z que encontramos.

Consecuentemente, la mezcla queda:

$$f_{X_{\text{MEZCLA}}} = \frac{9}{41} f_{X|X<4}(x) + \frac{32}{41} f_z(z)$$

Recordemos que el hecho de llamarlas Z fue solamente una cuestión de notación.

Los puntos que dividen las ramas de las funciones de densidad involucradas en la mezcla son: 1, 4, 2, 3.

Ahora estudiamos intervalo por intervalo:

- $-\infty < x < 1$

$$f_{X_{\text{MEZCLA}}} = \frac{9}{41} 0 + \frac{32}{41} 0 = 0$$

- $1 < x < 2$

$$f_{X_{\text{MEZCLA}}} = \frac{9}{41} \frac{2}{9}(x-1) + \frac{32}{41} 0 = \frac{2}{41}(x-1)$$

- $2 < x < 3$

$$f_{X_{\text{MEZCLA}}} = \frac{9}{41} \frac{2}{9}(x-1) + \frac{32}{41} \frac{1}{4}(2x-1) = \frac{2}{41}(9x-5)$$

- $3 < x < 4$

$$f_{X_{\text{MEZCLA}}} = \frac{9}{41} \frac{2}{9}(x-1) + \frac{32}{41} 0 = \frac{2}{41}(x-1)$$

- $4 < x < +\infty$

$$f_{X_{\text{MEZCLA}}} = \frac{9}{41} 0 + \frac{32}{41} 0 = 0$$

Ahora armamos la función de densidad de la mezcla, y queda:

$$f_{X_{\text{MEZCLA}}}(x) = \begin{cases} \frac{2}{41}(x-1) & 1 < x < 2 \\ \frac{2}{41}(9x-5) & 2 < x < 3 \\ \frac{2}{41}(x-1) & 3 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

b) Ahora debemos condicionar la distribución que obtuvimos en el punto a) para X

$= 2 \pm 0,1$. Es decir, para $1,9 < X < 2,1$.

$P(\text{seleccionar una varilla}) = P(1,9 < X < 2,1) =$

$$= \int_{1,9}^{2,1} f_{\text{XMEZCLA}} dx = \int_{1,9}^2 \frac{2}{41} (x-1) dx + \int_2^{2,1} \frac{2}{41} (9x-5) dx = 0,0702$$

$$f_{X/1,9 < X < 2,1}(x) = \begin{cases} \frac{1}{0,0702} \frac{2}{41} (x-1) & 1,9 < x < 2 \\ \frac{1}{0,0702} \frac{2}{41} (9x-5) & 2 < x < 2,1 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} 0,6944 (x-1) & 1,9 < x < 2 \\ 0,6944 (9x-5) & 2 < x < 2,1 \\ 0 & \forall \text{ otro } x \end{cases}$$

Y esa es la distribución de las varillas que selecciona.



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/process/text/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 13 de mayo de 2004

CAPÍTULO III

Variable Aleatoria Bidimensional y n-Dimensional

A veces es necesario trabajar con probabilidades que involucran a más de una variable aleatoria al mismo tiempo. Por ejemplo, podemos querer calcular cuál es la probabilidad de que una persona elegida al azar mida entre 1.70 y 1.80m y pese entre 80 y 90kg. O la probabilidad de que una persona que pesa entre 70 y 80kg mida menos de 1.60m.

En esos casos usamos variables aleatorias bidimensionales. En general, las variables aleatorias pueden ser de dimensión n ; hablamos entonces de variables n -dimensionales.

Las variables aleatorias que estudiamos en el capítulo anterior son un caso particular, denominado variable aleatoria unidimensional. En general, podemos pensar a las variables aleatorias n -dimensionales como vectores, siendo cada una de las componentes del vector una variable aleatoria en sí.

En el caso de las variables aleatorias bidimensionales, podemos pensarlas como un vector con dos componentes, cada una de las cuales es una variable aleatoria unidimensional tal cual las estudiamos en el capítulo anterior.

Así como los valores posibles de una variable unidimensional están contenidos en una recta, siendo la recta misma, o parte de ella, los valores posibles de una variable

bidimensional están contenidos en un plano, siendo todo el plano o parte de él.

Las variables aleatorias unidimensionales pueden ser discretas o continuas. Como cada componente de una variable aleatoria de dimensión mayor a 1 es una variable aleatoria unidimensional, una variable aleatoria bidimensional puede tener sus dos componentes discretas, sus dos componentes continuas, o una discreta y una continua.

Ejemplo

Se tiene el experimento "tirar un dado y una moneda"

El espacio muestral es $E = \{1\text{-cara}, 2\text{-cara}, 3\text{-cara}, 4\text{-cara}, 5\text{-cara}, 6\text{-cara}, 1\text{-ceca}, 2\text{-ceca}, 3\text{-ceca}, 4\text{-ceca}, 5\text{-ceca}, 6\text{-ceca}\}$

Definiremos una variable aleatoria bidimensional agrupando estas dos:

X: el número que sale al tirar el dado

Y: la cantidad de caras que salen

A raíz de la forma en que hemos tomado las variables, podríamos reescribir el espacio muestral así:

$E = \{ (1,0), (2,0), (3,0), (4,0), (5,0), (6,0), (1,1), (2,1), (3,1), (4,1), (5,1), (6,1) \}$

Si luego nos preguntan cuál es la probabilidad de que sacar un 5 en el dado y sacar cara en la moneda, es decir, $P(X = 5 \wedge Y = 1)$, pensaremos en que "X=5" y "Y=1" son dos sucesos independientes, y entonces $P(X = 5 \wedge Y = 1) = P(X = 5) P(Y = 1) = 1/6 \cdot 1/2 = 1/12$.

También podríamos pensar que los 12 resultados posibles de E son equiprobables, y entonces cada resultado tiene probabilidad 1/12.

En este ejemplo, las dos componentes de la variable aleatoria bidimensional son discretas. El ejemplo que dimos al principio, del peso y la altura de una persona, tiene sus dos componentes continuas. Un ejemplo de una variable aleatoria bidimensional con una componente discreta y una continua, puede ser considerar la longitud de las rutas y la cantidad de estaciones de servicio que hay en ellas.



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/process/text/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 13 de mayo de 2004

Distribución de probabilidad conjunta

Así como en las variables aleatorias unidimensionales nos interesa estudiar cómo se distribuye la probabilidad de cada uno de los valores posibles, en las variables aleatorias bidimensionales nos interesa lo mismo, con la salvedad de que ahora los valores posibles son pares de valores, o bien vectores de dimensión 2.

Notemos que:

- 1) la probabilidad de un determinado par de valores no puede ser menor que cero.
- 2) la suma de las probabilidades de todos los pares de valores da 1, porque al hacer el experimento siempre sale uno de los pares posibles.

Función de densidad de probabilidad conjunta

En las variables aleatorias unidimensionales, la función de densidad de probabilidad es una función que le asigna a cada valor posible de la variable aleatoria un número real que consiste en la probabilidad de que ocurra. En las variables aleatorias bidimensionales, la imagen de la función sigue siendo de dimensión 1 (porque la probabilidad es un número) pero el dominio es de dimensión 2.

Si X e Y son discretas:

$P_{XY}(x,y)$ es una función que a cada par de valores posibles le asigna su probabilidad.

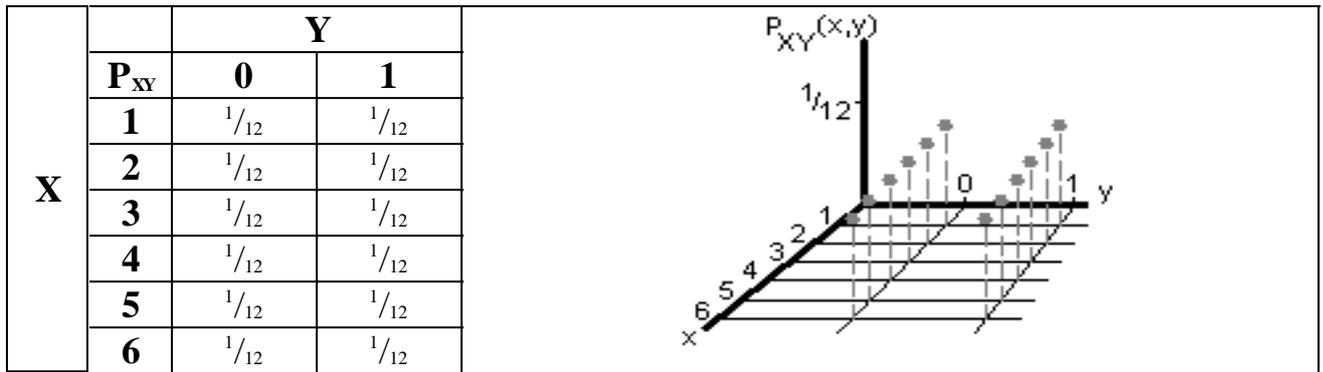
$P_{XY}(x,y)$ es una función de densidad de probabilidad discreta conjunta si y solo si cumple con:

- 1) $P_{XY}(x,y) \geq 0 \quad \forall (x,y)$
 $\sum_x \sum_y P_{XY}(x,y) = 1$
- 2)

Ejemplo:

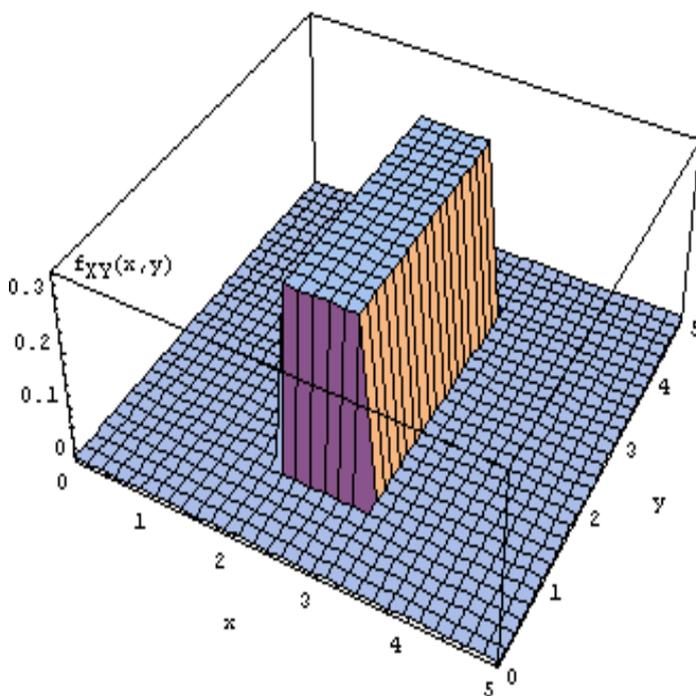
X: el número que sale al tirar un dado honesto.

Y: la cantidad de caras que salen al tirar una moneda.



Si X e Y son continuas:

Análogamente a la función de densidad de una variable aleatoria unidimensional, para obtener probabilidades a partir de la función de densidad de una variable aleatoria bidimensional debemos integrarla. En vez de una integral simple, es una integral doble. Es decir, la integral de la función de densidad $f_{XY}(x,y)$ es un dominio D del plano XY, da la probabilidad de que la variable aleatoria XY asuma un valor comprendido en ese dominio.



$f_{XY}(x,y)$ es una función de densidad de probabilidad continua conjunta si y solo si cumple con:

1) $f_{XY}(x,y) \geq 0 \quad \forall (x,y)$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x,y) dy dx = 1$$

2) $-\infty-\infty$

Ejemplo:

Se toma un punto al azar del plano XY, con la primera componente entre 2 y 3, y la segunda entre 1 y 4, y se toma la variable aleatoria X como la componente X del punto, y la variable aleatoria Y como la componente Y del punto.

$$f_{XY}(x, y) = \begin{cases} \frac{1}{3} & 2 < x < 3, 1 < y < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

Luego la probabilidad de que el par (X,Y) caiga en un determinado intervalo es la integral de la altura f_{XY} en dicho intervalo.

Problemas típicos

1) Determine si las siguientes funciones son de densidad de probabilidad conjunta:

a)

P_{XY}	Y			
	0	1	2	
X	1	0,2	0,1	0,1
	2	0,1	0,2	0,3
	3	0,2	0,1	0,2

b)

P_{XY}	Y		
	1	2	3
	0,3	-0,1	0,2
	0,4	0,1	0,1

c)

P_{XY}	Y		
	20	30	
X	1	0,1	0,3
	2	0,4	0,2

Resolución

$$\sum_x \sum_y P_{XY}(x, y) \neq 1$$

a) No, porque

b) No, porque $\exists (x,y)$ tal que $P_{XY}(x,y) < 0$

c) Sí.

2) Dada la siguiente distribución de probabilidad conjunta:

P_{XY}	Y			
	1	2	3	5

X	0	5/35	3/35	2/35	2/35
	1	2/35	4/35	3/35	1/35
	2	2/35	1/35	2/35	2/35
	3	1/35	1/35	2/35	2/35

Calcule

- a) $P(X=1 \wedge Y=2)$
- b) $P(X=Y)$
- c) $P(X<Y)$
- d) $P(X=1)$

Resolución

a) Directamente de la tabla, $P(X=1 \wedge Y=2) = 4/35$

b) La probabilidad de un determinado suceso es la suma de las probabilidades de los sucesos simples que lo forman. En relación a la tabla, es la suma de las probabilidades de los casilleros que cumplen con la condición. Entonces:

$$P(X=Y) = P(X=1 \wedge Y=1) + P(X=2 \wedge Y=2) + P(X=3 \wedge Y=3) = 2/35 + 1/35 + 2/35 = 5/35$$

$$\begin{aligned} c) P(X<Y) &= P(X=0 \wedge Y=1) + P(X=0 \wedge Y=2) + P(X=0 \wedge Y=3) + P(X=0 \wedge Y=5) \\ &+ P(X=1 \wedge Y=2) + P(X=1 \wedge Y=3) + P(X=1 \wedge Y=5) + P(X=2 \wedge Y=3) + P(X=2 \wedge Y=5) \\ &+ P(X=3 \wedge Y=5) = 5/35 + 3/35 + 2/35 + 2/35 + 4/35 + 3/35 + 1/35 + 2/35 + 2/35 + 2/35 = 26/35 \end{aligned}$$

Aunque habría sido más económico calcular:

$$\begin{aligned} P(X<Y) &= 1 - P(X \geq Y) = P(X=1 \wedge Y=1) + P(X=2 \wedge Y=1) + P(X=2 \wedge Y=2) + \\ &P(X=3 \wedge Y=1) + P(X=3 \wedge Y=2) + P(X=3 \wedge Y=3) = 2/35 + 2/35 + 1/35 + 1/35 + 1/35 + 2/35 = 9/35 \end{aligned}$$

$$\begin{aligned} d) P(X=1) &= P(X=1 \wedge Y=1) + P(X=1 \wedge Y=2) + P(X=1 \wedge Y=3) + P(X=1 \wedge Y=5) \\ &= 2/35 + 4/35 + 3/35 + 1/35 = 10/35 \end{aligned}$$

3) Determine si las siguientes funciones son de densidad de probabilidad conjunta:

a)

$$f_{XY}(x, y) = \begin{cases} \frac{3}{2} \left((x+y)^2 - \frac{1}{2} \right) & -1 < x < 1, -1 < y < 1 \\ 0 & \forall \text{ otro } (x, y) \end{cases}$$

b)

$$f_{XY}(x, y) = \begin{cases} \frac{3}{8} (x+y)^2 & -1 < x < 1, -1 < y < 1 \\ 0 & \forall \text{ otro } (x, y) \end{cases}$$

c)

$$f_{XY}(x, y) = \begin{cases} \frac{3}{4} (x+y)^2 & -1 < x < 1, -1 < y < 1 \\ 0 & \forall \text{ otro } (x, y) \end{cases}$$

Resolución

- a) No, porque $\exists (x,y)$ tal que $f_{XY}(x,y) < 0$
 b) Sí.

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx \neq 1$$

- c) No, porque

4) Hallar k para que $f_{XY}(x,y)$ sea función de densidad:

$$f_{XY}(x, y) = \begin{cases} k(x^2 + y) & 0 < x < 2, 0 < y < 3 \\ 0 & \forall \text{ otro } (x, y) \end{cases}$$

a)

$$f_{XY}(x, y) = \begin{cases} k(x + y) & 1 < x < 2, 0 < y < 1 \\ 2x - y & 1 < x < 2, 1 < y < 2 \\ 0 & \forall \text{ otro } (x, y) \end{cases}$$

b)

Resolución

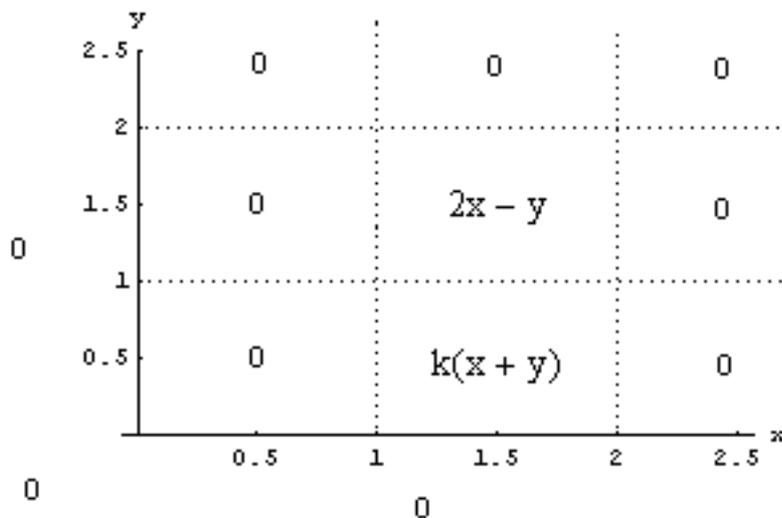
a) Sabemos que la integral sobre el plano XY de la función de densidad conjunta debe cerrar a 1, entonces:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx = 1$$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx = \int_0^3 \int_0^2 k(x^2 + y) dy dx = \dots = 17k = 1$$

Luego $k = 1/17$

b) En esta caso hacemos lo mismo, pero la función es un poco más compleja porque tiene más de una rama. Hacemos el gráfico para orientarnos:



$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx = \int_1^2 \int_0^1 k(x + y) dy dx + \int_1^2 \int_1^2 (2x - y) dy dx = \dots = 2k + \frac{1}{4} = 1$$

Luego $k = 3/8$

5) Calcular las siguientes probabilidades:

- a) $P(X > 1 \wedge Y < 2)$, con la $f_{XY}(x,y)$ del ejercicio 4.a.
- b) $P(X > 3/2 \wedge Y > 1/2)$, con la $f_{XY}(x,y)$ del ejercicio 4.b.

$$f_{XY}(x, y) = \begin{cases} e^{-(x+y)} & x > 0, y > 0 \\ 0 & \forall \text{ otro } (x, y) \end{cases}$$

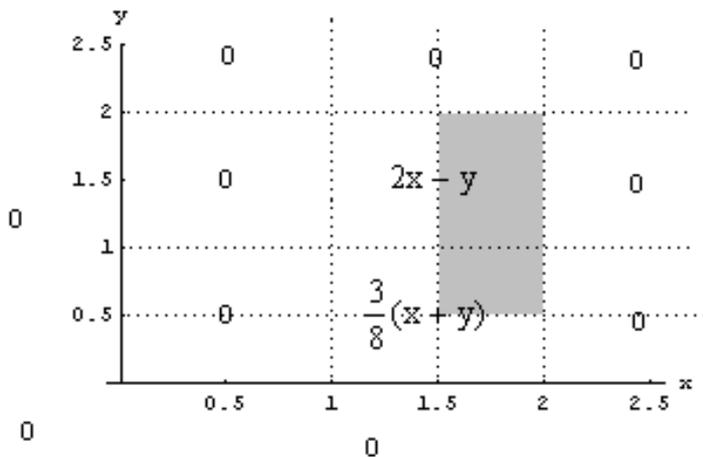
- c) $P(X > 2 \wedge Y < 1)$, con

Resolución

$$P(X > 1 \wedge Y < 2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx = \int_1^{+\infty} \int_0^2 \frac{x^2 + y}{17} dy dx = 0,392157$$

a)

b) Hacemos un diagrama para no equivocarnos con la integral:



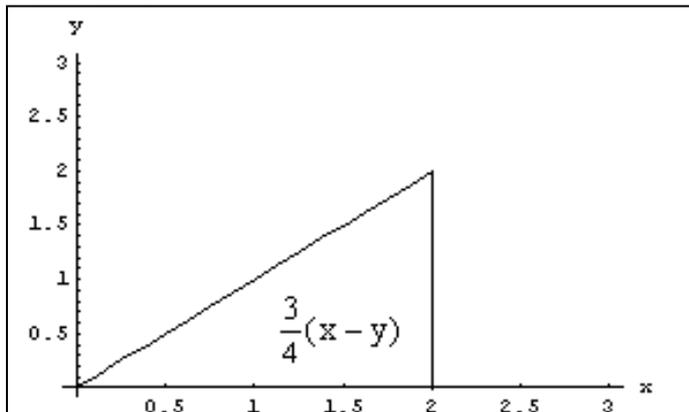
$$\begin{aligned} P(X > \frac{3}{2} \wedge Y > \frac{1}{2}) &= \int_{3/2}^{+\infty} \int_{1/2}^{+\infty} f_{XY}(x, y) dy dx = \\ &= \int_{3/2}^{+\infty} \int_{1/2}^2 \frac{3}{8}(x+y) dy dx + \int_{3/2}^{+\infty} \int_1^{+\infty} (2x-y) dy dx = \\ &= \dots = 0,2344 + \frac{1}{6} = 0,401 \end{aligned}$$

6) Dada la siguiente función de densidad:

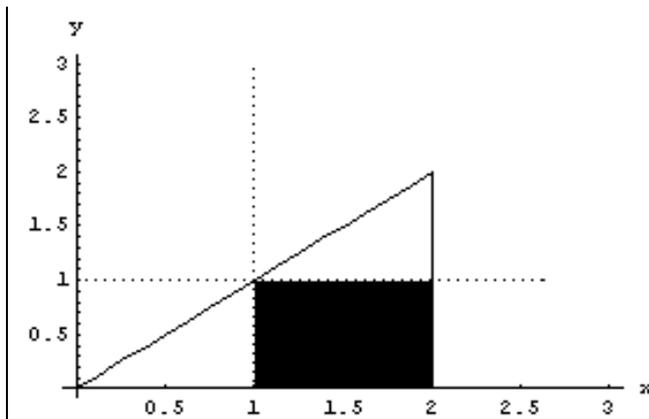
$$f_{XY}(x, y) = \begin{cases} \frac{3}{4}(x-y) & 0 < x < 2, 0 < y < x \\ 0 & \forall \text{ otro } (x, y) \end{cases}$$

- a) Calcular $P(X > 1 \wedge Y < 1)$
- b) Calcular $P(X > 1)$
- c) Calcular $P(Y < 1)$

Resolución



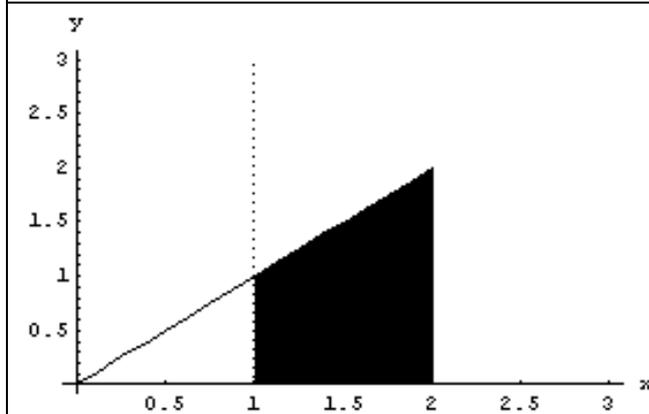
La distribución conjunta con la que vamos a trabajar



a)

$$P(X > 1 \wedge Y < 1) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx =$$

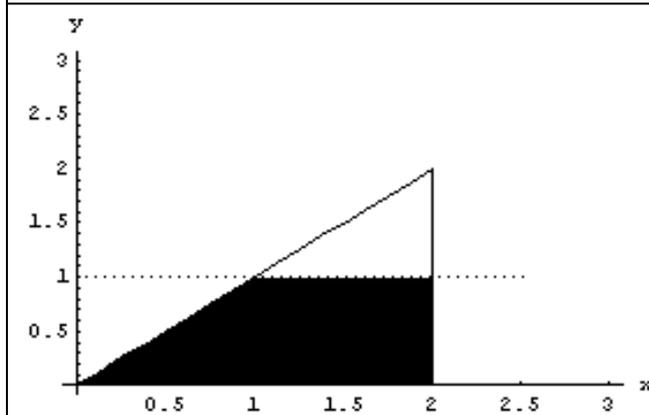
$$\int_0^2 \int_0^1 \frac{1}{4}(x-y) dy dx = \dots = \frac{3}{4}$$



b)

$$P(X > 1) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx =$$

$$\int_0^2 \int_0^3 \frac{1}{4}(x-y) dy dx = \dots = \frac{7}{8}$$



c)

$$P(Y < 1) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy =$$

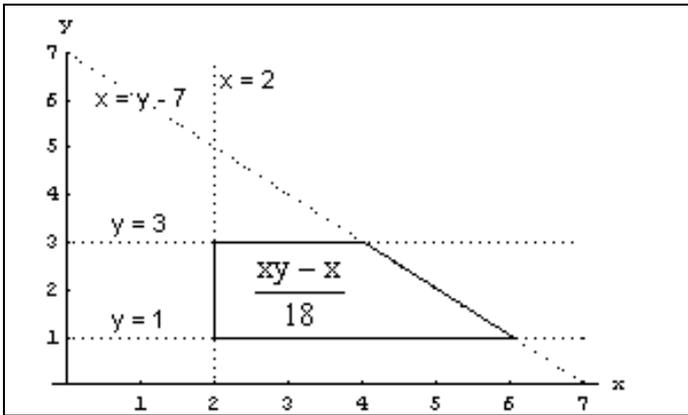
$$\int_0^1 \int_0^2 \frac{1}{4}(x-y) dx dy = \dots = \frac{7}{8}$$

7) Dada la siguiente función de densidad:

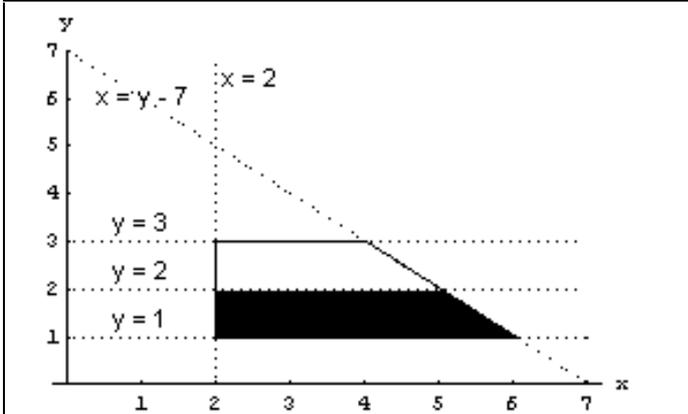
$$f_{XY}(x, y) = \begin{cases} \frac{xy - x}{18} & 1 < y < 3, 2 < x < 7 - y \\ 0 & \forall \text{ otro } (x, y) \end{cases}$$

- a) Calcular $P(Y < 2)$
- b) Calcular $P(X > 3)$
- c) Calcular $P(X > 3 \wedge Y > 2)$

Resolución



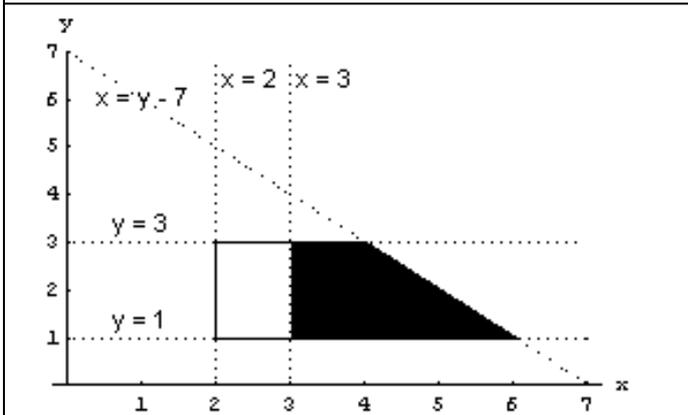
La distribución conjunta con la que vamos a trabajar



a)

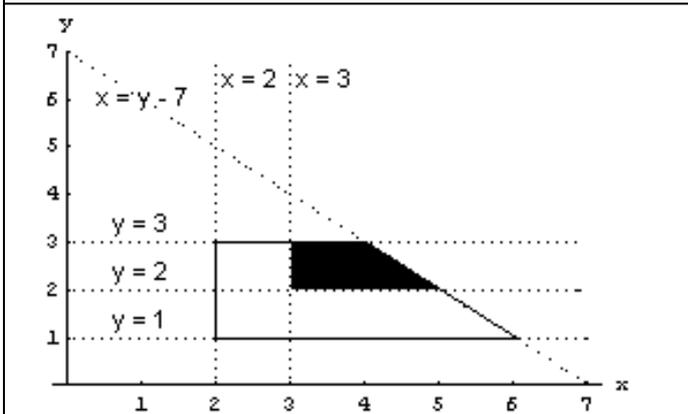
$$P(Y < 2) = \int_{-\infty}^{-\infty} \int_{-\infty}^{-\infty} f_{XY}(x, y) dx dy =$$

$$\int_1^2 \int_2^{7-y} \frac{xy-x}{18} dx dy = \dots = 0,340278$$



b)

$$P(X > 3) = \int_1^3 \int_3^{7-y} \frac{xy-x}{18} dx dy = \dots = 0,722222$$



c)

$$P(X > 3 \wedge Y > 2) = \int_2^3 \int_3^{7-y} \frac{xy-x}{18} dx dy = \dots = 0,451389$$



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/processtext/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 13 de mayo de 2004_

Distribuciones marginales

Como vimos antes, cada componente de una variable aleatoria bidimensional es una variable aleatoria unidimensional en sí misma. Es decir, cada una de las dos variables aleatorias que forman la variable aleatoria bidimensional es una variable aleatoria unidimensional común y corriente. Entonces nos puede interesar conocer la distribución de una componente por separado, sin tener en cuenta a la otra componente.

Eso se denomina "marginal", y la distribución de la variable unidimensional por separado se llama "distribución marginal".

Distribuciones marginales de variables aleatorias discretas

Sea la variable aleatoria bidimensional XY distribuida según $P_{XY}(x,y)$, la distribución de X (también llamada distribución marginal de X) es:

$$P_X(x) = \sum_{y=-\infty}^{\infty} P_{XY}(x,y)$$

para cada valor x de la variable aleatoria X

Análogamente, la distribución de Y es:

$$P_Y(y) = \sum_{x=-\infty}^{\infty} P_{XY}(x,y)$$

para cada valor y de la variable aleatoria Y

Es decir, para cada valor posible de la variable aleatoria cuya distribución se desea hallar, se suman las probabilidades conjuntas de ese valor con cada uno de los valores posibles de la otra variable.

Ejemplo 1

Si la distribución conjunta es:

P_{xy}		Y	
		20	30
X	1	0,1	0,3
	2	0,4	0,2

Vamos a hallar la distribución de X.

Primero enumeramos los valores posibles de X: 1; 2.

Y ahora para cada valor posible de X, aplicamos la fórmula.

$$P_X(1) = \sum_{y=-\infty} P_{XY}(1, y) = P_{XY}(1, 20) + P_{XY}(1, 30) = 0,1 + 0,3 = 0,4$$

$$P_X(2) = \sum_{y=-\infty} P_{XY}(2, y) = P_{XY}(2, 20) + P_{XY}(2, 30) = 0,4 + 0,2 = 0,6$$

Entonces obtuvimos:

$$P_X(x) = \begin{cases} 0,4 & x = 1 \\ 0,6 & x = 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

Ahora hallemos la distribución de Y:

Primero enumeramos los valores posibles de Y: 20; 30.

Y ahora para cada valor posible de X, aplicamos la fórmula.

$$P_Y(20) = \sum_{x=-\infty} P_{XY}(x, 20) = P_{XY}(1, 20) + P_{XY}(2, 20) = 0,1 + 0,4 = 0,5$$

$$P_Y(30) = \sum_{x=-\infty} P_{XY}(x, 30) = P_{XY}(1, 30) + P_{XY}(2, 30) = 0,3 + 0,2 = 0,5$$

Entonces obtuvimos:

$$P_Y(y) = \begin{cases} 0,5 & y = 20 \\ 0,5 & y = 30 \\ 0 & \forall \text{ otro } y \end{cases}$$

Veamos lo que ocurre si en la tabla que usamos para escribir la distribución conjunta, agregamos los totales por fila y por columna:

P_{xy}		Y		
		20	30	
X	1	0,1	0,3	0,4
	2	0,4	0,2	0,6
		0,5	0,5	

Observamos que en los márgenes de la tabla no obtuvimos otra cosa que las distribuciones marginales de X y de Y. Esa es la razón por la cual las distribuciones de X e Y por separado se denominan "marginales".

Ejemplo 2

Vamos a hallar rápidamente las distribuciones de las variables X e Y, cuya distribución conjunta es la siguiente:

P_{XY}	Y				
	1	2	3	5	
X	0	5/35	3/35	2/35	2/35
	1	2/35	4/35	3/35	1/35
	2	2/35	1/35	2/35	2/35
	3	1/35	1/35	2/35	2/35

Como hicimos antes, anotaremos en los márgenes de la tabla los totales por fila y por columna:

P_{XY}	Y					
	1	2	3	5		
X	0	5/35	3/35	2/35	2/35	12/35
	1	2/35	4/35	3/35	1/35	10/35
	2	2/35	1/35	2/35	2/35	7/35
	3	1/35	1/35	2/35	2/35	6/35
	10/35	9/35	9/35	7/35		

$$P_X(x) = \left\{ \begin{array}{l} 12/35 \quad x=0 \\ 10/35 \quad x=1 \\ 7/35 \quad x=2 \\ 6/35 \quad x=3 \\ 0 \quad \forall \text{ otro } x \end{array} \right\} \quad P_Y(y) = \left\{ \begin{array}{l} 10/35 \quad y=1 \\ 9/35 \quad y=2 \\ 9/35 \quad y=3 \\ 7/35 \quad y=5 \\ 0 \quad \forall \text{ otro } y \end{array} \right\}$$

Luego

Distribuciones marginales de variables aleatorias continuas

La marginación de variables continuas es análoga a la de las variables discretas, pero puede acarrear algunas dificultades adicionales. Sea la variable aleatoria bidimensional XY distribuida según $f_{XY}(x,y)$, la distribución de X (también llamada distribución marginal de X) es:

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

para cada región del dominio de X donde no cambien los límites de integración de $f_{XY}(x,y)$ con respecto a Y.

Análogamente, la distribución de Y es:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y) dx$$

para cada región del dominio de Y donde no cambien los límites de integración de $f_{XY}(x,y)$ con respecto a X.

Es importante tener en cuenta las distintas ramas de $f_{XY}(x,y)$.

Método para obtener $f_X(x)$ a partir de $f_{XY}(x,y)$

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

Si bien la expresión $f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y) dy$ contiene toda la información necesaria para saber cómo obtener $f_X(x)$ a partir de $f_{XY}(x,y)$ sin importar cómo sea $f_{XY}(x,y)$, y la resolución de la integral no tiene nada que ver con probabilidad y estadística sino que constituye un tema de análisis matemático, a veces puede tornarse complicado, y convertirse en un verdadero impedimento **matemático** para quien intenta trabajar con la **estadística**.

Con temas similares se ofrece un método para resolverlos, pero en este caso es difícil establecer un método práctico y detallado que permita resolver los problemas mecánicamente. Por eso para este tema se ofrece un método sintetizado y una abundante cantidad de ejemplos.

Método:

- 1) Subdividir el dominio de X de forma tal que en cada intervalo no cambien:
 - Las ecuaciones que determinan los límites de integración de $f_{XY}(x,y)$ respecto de Y.
 - Las ecuaciones que determinan la separación de las ramas de $f_{XY}(x,y)$ (si las hay).

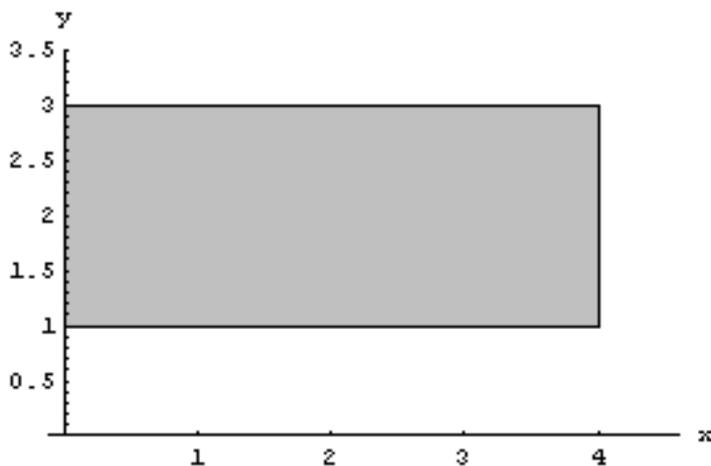
$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y) dy$$

2) Para cada intervalo, calcular $f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y) dy$, teniendo en cuenta que si en ese intervalo de X hay distintas ramas de $f_{XY}(x,y)$, la integral será la suma de distintas integrales.

3) Armar la $f_X(x)$ poniendo en cada intervalo lo calculado en el punto 2.

A continuación presentamos 22 ejemplos resueltos de marginación de variables.

Ejemplos 1, 2



$$f_{XY}(x, y) = \begin{cases} \frac{1}{8} & 0 < x < 4, 1 < y < 3 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Marginación de x:

Tenemos que subdividir el dominio de la X de forma tal que dentro de cada intervalo no cambien las ecuaciones que determinan los límites de integración respecto de Y, ni las que separan ramas de $f_{XY}(x, y)$.

En esta $f_{XY}(x, y)$ no hay múltiples ramas, así que para dividir en intervalos el dominio de X, solamente tendremos en cuenta el comportamiento de la Y en cada intervalo:

Para $-\infty < x < 0$, la Y no aparece.

Para $0 < x < 4$, la Y varía entre 1 y 3.

Para $4 < x < +\infty$, la Y no aparece.

Entonces nos quedan 3 intervalos, de los cuales 2 son triviales porque son imposibles (con lo cual sabemos que la densidad marginal de la X será cero en ellos).

Entonces aplicamos la fórmula al único intervalo relevante ($0 < x < 4$):

$$\int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \int_1^3 \frac{1}{8} dy = \frac{1}{4}$$

Luego construimos la función de densidad de la X, que tendrá solamente una rama porque hubo un solo intervalo relevante:

$$f_X(x) = \begin{cases} 1/4 & 0 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

Marginación de y:

En este caso la marginación de Y es muy similar a la de X. Como no hay múltiples ramas, solo vamos a observar el comportamiento de la X a la hora de tomar intervalos para la Y. Procedemos:

Para $-\infty < y < 1$, la X no aparece.

Para $1 < y < 3$, la X varía entre 0 y 4.

Para $3 < y < +\infty$, la X no aparece.

Entonces nos quedan 3 intervalos, de los cuales 2 son triviales porque son imposibles (con lo cual sabemos que la densidad marginal de la Y será cero en ellos).

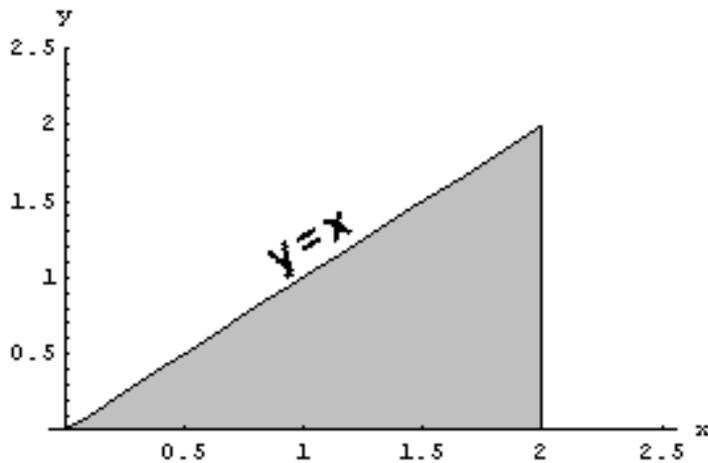
Entonces aplicamos la fórmula al único intervalo relevante ($1 < y < 3$):

$$\int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \int_0^4 \frac{1}{8} dx = \frac{1}{2}$$

Luego construimos la función de densidad de la Y, que tendrá solamente una rama porque hubo un solo intervalo relevante:

$$f_Y(y) = \begin{cases} 1/2 & 1 < y < 3 \\ 0 & \forall \text{ otro } y \end{cases}$$

Ejemplos 3, 4



$$f_{XY}(x, y) = \begin{cases} \frac{x+y}{4} & 0 < x < 2, 0 < y < x \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Marginación de x:

Como la $f_{XY}(x, y)$ tiene una sola rama, vamos a tener en cuenta solamente el comportamiento de Y:

Para $-\infty < x < 0$, la Y no aparece.

Para $0 < x < 2$, la Y varía entre 0 y x.

Para $2 < x < +\infty$, la Y no aparece.

Entonces nos quedan 3 intervalos, de los cuales 2 son triviales porque son imposibles (con lo cual sabemos que la densidad marginal de la X será cero en ellos).

Entonces aplicamos la fórmula al único intervalo relevante ($0 < x < 2$):

$$\int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \int_0^x \frac{x+y}{4} dy = \frac{3x^2}{8}$$

Luego construimos la función de densidad de la X, que tendrá solamente una rama porque hubo un solo intervalo relevante:

$$f_X(x) = \begin{cases} \frac{3x^2}{8} & 0 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

Marginación de y:

Como la $f_{XY}(x, y)$ tiene una sola rama, vamos a tener en cuenta solamente el comportamiento de X:

Para $-\infty < y < 0$, la X no aparece.

Para $0 < y < 2$, la X varía entre y y 2.

Para $2 < y < +\infty$, la X no aparece.

Entonces nos quedan 3 intervalos, de los cuales 2 son triviales porque son imposibles (con lo cual sabemos que la densidad marginal de la Y será cero en ellos).

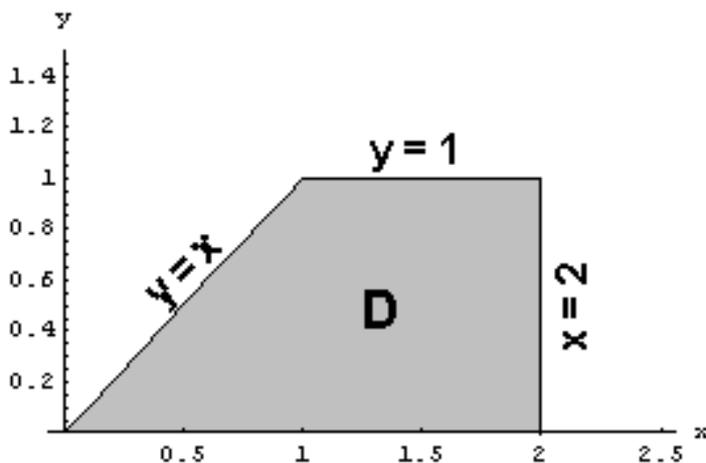
Entonces aplicamos la fórmula al único intervalo relevante ($0 < y < 2$):

$$\int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \int_y^2 \frac{x+y}{4} dx = \frac{4+4y-3y^2}{8}$$

Luego construimos la función de densidad de la Y, que tendrá solamente una rama porque hubo un solo intervalo relevante:

$$f_Y(y) = \begin{cases} \frac{4+4y-3y^2}{8} & 0 < y < 2 \\ 0 & \forall \text{ otro } y \end{cases}$$

Ejemplos 5, 6



$$f_{XY}(x, y) = \begin{cases} \frac{8xy}{7} & (x, y) \in D \\ 0 & (x, y) \notin D \end{cases}$$

donde D es el que se ve en el gráfico.

Marginación de x:

Como la $f_{XY}(x,y)$ tiene una sola rama, vamos a tener en cuenta solamente el comportamiento de Y:

Para $-\infty < x < 0$, la Y no aparece.

Para $0 < x < 1$, la Y varía entre 0 y x.

Para $1 < x < 2$, la Y varía entre 0 y 1.

Para $2 < x < +\infty$, la Y no aparece.

Entonces nos quedan 4 intervalos, de los cuales 2 son triviales porque son imposibles (con lo cual sabemos que la densidad marginal de la X será cero en ellos).

Entonces aplicamos la fórmula a los intervalos relevantes:

- $0 < x < 1$

$$\int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \int_0^1 \frac{8xy}{7} dy = \frac{4x^3}{7}$$

- $1 < x < 2$

$$\int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \int_0^1 \frac{8xy}{7} dy = \frac{4x}{7}$$

Luego construimos la función de densidad de la X, que tendrá 2 ramas porque hubo dos intervalos relevantes:

$$f_X(x) = \begin{cases} \frac{4x^3}{7} & 0 < x < 1 \\ \frac{4x}{7} & 1 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

Marginación de y:

Como la $f_{XY}(x, y)$ tiene una sola rama, vamos a tener en cuenta solamente el comportamiento de X:

Para $-\infty < y < 0$, la X no aparece.

Para $0 < y < 1$, la X varía entre y y 2.

Para $1 < y < +\infty$, la X no aparece.

Entonces nos quedan 3 intervalos, de los cuales 2 son triviales porque son imposibles (con lo cual sabemos que la densidad marginal de la Y será cero en ellos).

Entonces aplicamos la fórmula al único intervalo relevante ($0 < y < 1$):

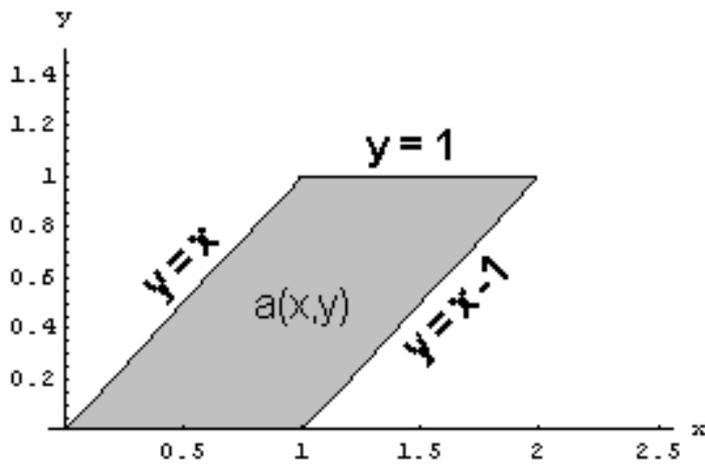
$$\int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \int_y^2 \frac{8xy}{7} dx = \frac{4}{7} y(4 - y^2)$$

Luego construimos la función de densidad de la Y, que tendrá solamente una rama porque hubo un solo intervalo relevante:

$$f_Y(y) = \begin{cases} \frac{4}{7} y(4 - y^2) & 0 < y < 1 \\ 0 & \forall \text{ otro } y \end{cases}$$

Notemos que no siempre las funciones de densidad marginales de X e Y tienen la misma cantidad de ramas. En este ejemplo observamos que la de la X tiene 2 mientras que la de la Y tiene solamente 1.

Ejemplos 7, 8



Supongamos que este es el gráfico de una función de densidad conjunta de X e Y, con una sola rama cuya expresión es una determinada función $a(x,y)$.

A partir de ahora no trabajaremos más con distribuciones concretas sino que lo haremos abstractamente para que no nos distraigan las cuentas.

Marginación de x:

Para $-\infty < x < 0$, la Y no aparece.

Para $0 < x < 1$, la Y varía entre 0 y x.

Para $1 < x < 2$, la Y varía entre $x-1$ y 1

Para $2 < x < +\infty$, la Y no aparece.

Aplicamos la fórmula a los intervalos relevantes, y la función de densidad marginal de X nos queda:

$$f_X(x) = \begin{cases} \int_0^x a(x, y) dy & 0 < x < 1 \\ \int_{x-1}^1 a(x, y) dy & 1 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

Marginación de y:

Para $-\infty < y < 0$, la X no aparece.

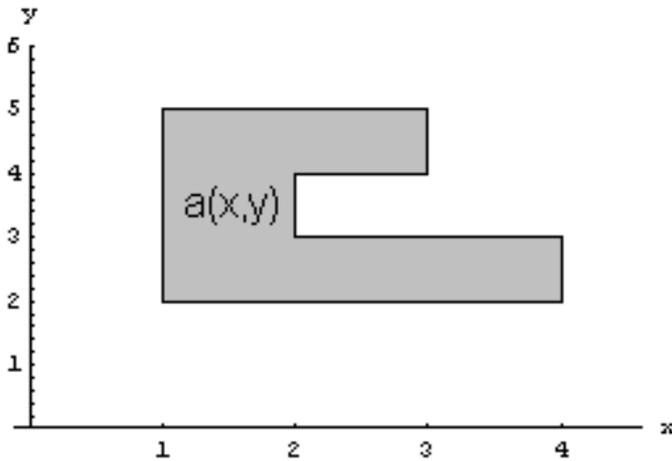
Para $0 < y < 1$, la X varía entre "y" y "y+1".

Para $1 < y < +\infty$, la X no aparece.

Aplicamos la fórmula al intervalo relevante, y la función de densidad marginal de Y nos queda:

$$f_Y(y) = \begin{cases} \int_y^{y+1} a(x, y) dx & 0 < y < 1 \\ 0 & \forall \text{ otro } y \end{cases}$$

Ejemplos 9, 10



Supongamos que este es el gráfico de una función de densidad conjunta de X e Y, con una sola rama cuya expresión es una determinada función $a(x,y)$.

<p>Marginación de x:</p> <ul style="list-style-type: none"> • $-\infty < x < 1$: la Y no aparece. • $1 < x < 2$: la Y varía entre 2 y 5. • $2 < x < 3$: la Y varía entre 2 y 3, y también entre 4 y 5 • $3 < x < 4$: la Y varía entre 2 y 3 • $4 < x < +\infty$: la Y no aparece. 	$f_x(x) = \begin{cases} \int_2^5 a(x,y) dy & 1 < x < 2 \\ \int_2^3 a(x,y) dy + \int_4^5 a(x,y) dy & 2 < x < 3 \\ \int_2^3 a(x,y) dy & 3 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$
<p>Cuando en un intervalo de la X, la Y hace más de una aparición, se suman las integrales correspondientes a cada a aparición. Vemos en el ejemplo que en la rama $2 < x < 3$ de la $f_x(x)$, aparece la suma de las dos integrales correspondientes a las dos apariciones de la Y.</p>	

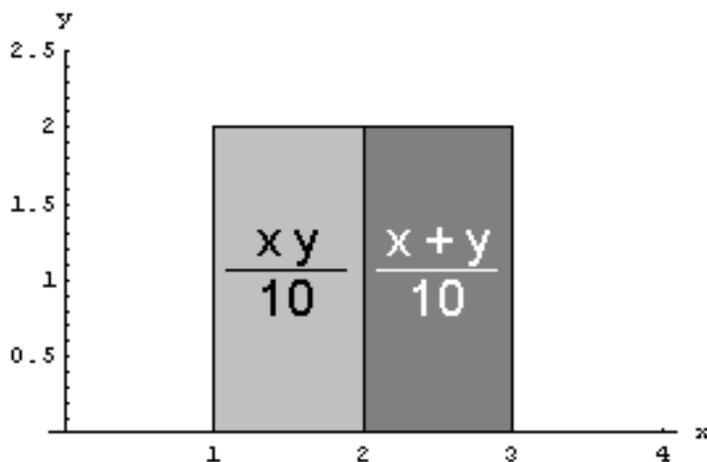
Marginación de y:

- $-\infty < y < 2$: la X no aparece.
- $2 < y < 3$: la X varía entre 1 y 4
- $3 < y < 4$: la X varía entre 1 y 2
- $4 < y < 5$: la X varía entre 1 y 3
- $5 < y < +\infty$: la X no aparece.

$$f_X(x) = \begin{cases} \int_2^4 a(x,y) dy & 2 < y < 3 \\ \int_1^2 a(x,y) dy & 3 < y < 4 \\ \int_1^3 a(x,y) dy & 4 < y < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

A partir de ahora veremos ejemplos en los que la $f_{XY}(x,y)$ tiene más de una rama.

Ejemplos 11, 12



$$f_{XY}(x,y) = \begin{cases} \frac{x \cdot y}{10} & 1 < x < 2, 0 < y < 2 \\ \frac{x + y}{10} & 2 < x < 3, 0 < y < 2 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Cuando la función de densidad conjunta tiene más de una rama, se procede como veníamos haciendo hasta ahora, con la diferencia de que en vez de estudiar las apariciones de la Y, estudiamos las apariciones de cada rama.

Llamemos para este ejemplo rama "a" a la de la izquierda y rama "b" a la de la derecha.

Marginación de x:

- $-\infty < x < 1$:
 - la rama a no aparece
 - la rama b no aparece

- $1 < x < 2$:
 - la rama a aparece entre 0 y 2
 - la rama b no aparece
- $2 < x < 3$:
 - la rama a no aparece
 - la rama b aparece entre 0 y 2
- $3 < x < +\infty$:
 - la rama a no aparece
 - la rama b no aparece

Nos quedaron 4 intervalos, pero el primero y el último son triviales porque en ellos no aparece nada. Con los intervalos relevantes procedemos como antes, pero ahora teniendo cuidado porque el integrando no siempre será el mismo en todas las integrales, sino que ahora dependerá de la rama que haya aparecido.

Entonces:

- $1 < x < 2$:

$$\int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \int_0^2 \frac{xy}{10} dy = \frac{x}{5}$$

- $2 < x < 3$:

$$\int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \int_0^2 \frac{x+y}{10} dy = \frac{x+1}{5}$$

Luego la función de densidad marginal de X es:

$$f_X(x) = \begin{cases} \frac{x}{5} & 1 < x < 2 \\ \frac{x+1}{5} & 2 < x < 3 \\ 0 & \forall \text{ otro } x \end{cases}$$

Marginación de y:

- $-\infty < y < 0$:
 - la rama a no aparece
 - la rama b no aparece
- $0 < y < 2$:
 - la rama a aparece entre 1 y 2
 - la rama b aparece entre 2 y 3
- $2 < y < +\infty$:
 - la rama a no aparece
 - la rama b no aparece

El único intervalo no trivial es $0 < y < 2$. Observamos que además en ese intervalo hay más de una aparición. Como puede intuirse, lo que se hace es sumar las

correspondientes integrales.

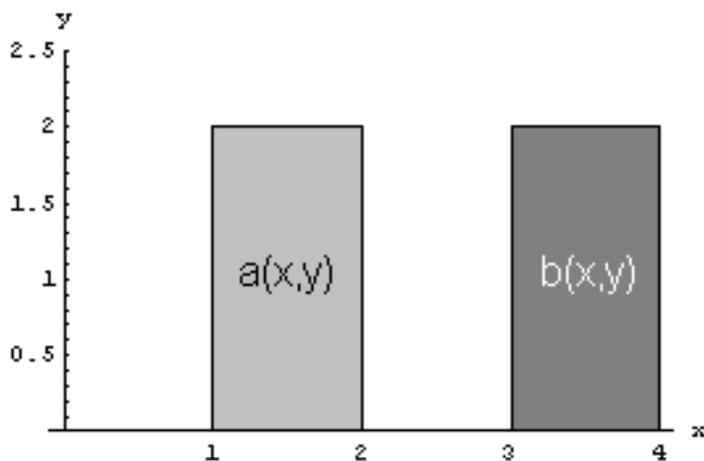
Entonces, aplicando la fórmula en $0 < y < 2$ obtenemos:

$$\int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \int_1^2 f_{XY}(x, y) dx = \int_1^2 f_{XY}(x, y) dx + \int_2^4 f_{XY}(x, y) dx = \int_1^2 \frac{xy}{10} dx + \int_2^4 \frac{x+y}{10} dx = \frac{y+1}{4}$$

Luego la función de densidad marginal de Y es:

$$f_Y(y) = \begin{cases} \frac{y+1}{4} & 0 < y < 2 \\ 0 & \forall \text{ otro } y \end{cases}$$

Ejemplos 13, 14

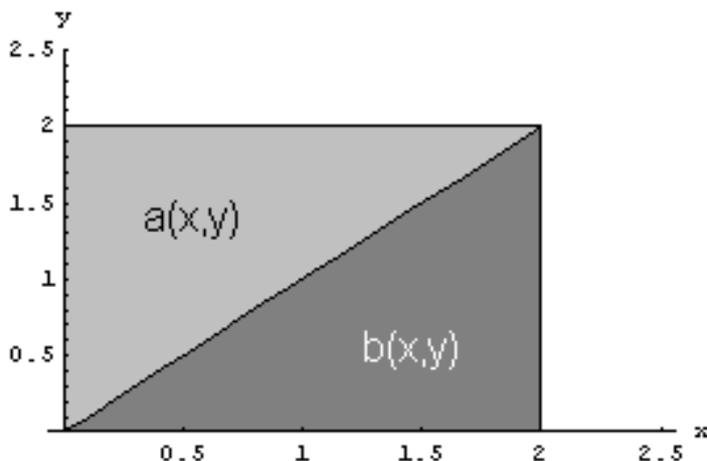


$$f_{XY}(x, y) = \begin{cases} a(x, y) & 1 < x < 2, 0 < y < 2 \\ b(x, y) & 3 < x < 4, 0 < y < 2 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Este caso es similar al anterior, por lo cual lo haremos rápidamente. Además, de aquí en adelante, trabajaremos con funciones de densidad genéricas (con letras) para no hacer cuentas que nos distraigan del objetivo primario.

<p>Marginación de x:</p> <ul style="list-style-type: none"> • $-\infty < x < 1$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b no aparece • $1 < x < 2$: <ul style="list-style-type: none"> • la rama a aparece entre 0 y 2 • la rama b no aparece • $2 < x < 3$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b no aparece • $3 < x < 4$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b aparece entre 0 y 2 • $4 < x < +\infty$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b no aparece 	<p>Se revuelven las siguientes integrales:</p> $\int_0^2 a(x, y) dy \quad ; \quad \int_0^2 b(x, y) dy$ <p>Y se obtiene:</p> $f_x(x) = \begin{cases} \int_0^2 a(x, y) dy & 1 < x < 2 \\ \int_0^2 b(x, y) dy & 3 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$
<p>Marginación de y:</p> <ul style="list-style-type: none"> • $-\infty < y < 0$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b no aparece • $0 < y < 2$: <ul style="list-style-type: none"> • la rama a aparece entre 1 y 2 • la rama b aparece entre 3 y 4 • $2 < y < +\infty$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b no aparece 	<p>Se revuelven las siguientes integrales:</p> $\int_1^2 a(x, y) dx \quad ; \quad \int_3^4 b(x, y) dx$ <p>Y se obtiene:</p> $f_y(y) = \begin{cases} \int_1^2 a(x, y) dx + \int_3^4 b(x, y) dx & 0 < y < 2 \\ 0 & \forall \text{ otro } y \end{cases}$

Ejemplos 15, 16



$$f_{XY}(x, y) = \begin{cases} a(x, y) & 0 < x < 2, x < y < 2 \\ b(x, y) & 0 < x < 2, 0 < y < x \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Marginación de x:

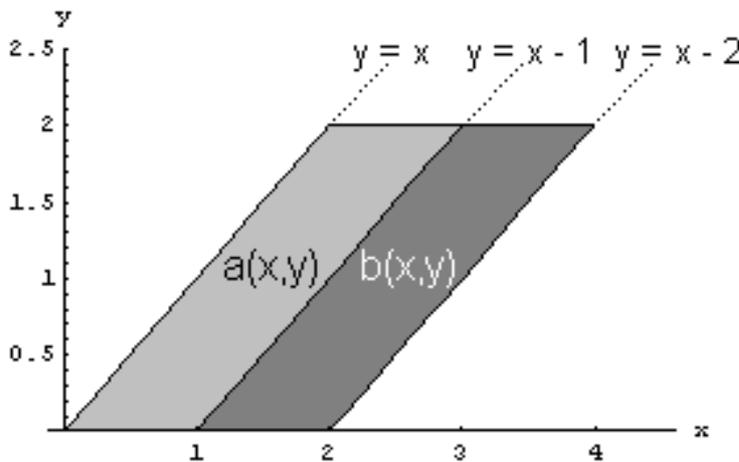
- $-\infty < x < 0$:
 - la rama a no aparece
 - la rama b no aparece
- $0 < x < 2$:
 - la rama a aparece entre **x** y **2**
 - la rama b aparece entre **0** y **x**
- $2 < x < +\infty$:
 - la rama a no aparece
 - la rama b no aparece

La integral a resolver es:

$$\int_0^2 f_{XY}(x, y) dy = \int_0^2 f_{XY}(x, y) dy + \int_x^2 f_{XY}(x, y) dy + \int_0^x b(x, y) dy + \int_x^2 a(x, y) dy$$

Se obtiene:

$$f_x(x) = \begin{cases} \int_0^x b(x, y) dy + \int_x^2 a(x, y) dy & 0 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases}$$



La integral a resolver es:

$$\int_0^2 f_{XY}(x, y) dx = \int_0^2 f_{XY}(x, y) dx + \int_y^2 f_{XY}(x, y) dx + \int_0^y a(x, y) dx + \int_y^2 b(x, y) dx$$

Se obtiene:

$$f_y(y) = \begin{cases} \int_0^y a(x, y) dx + \int_y^2 b(x, y) dx & 0 < y < 2 \\ 0 & \forall \text{ otro } y \end{cases}$$

Marginación de y:

- $-\infty < y < 0$:
 - la rama a no aparece
 - la rama b no aparece
- $0 < y < 2$:
 - la rama a aparece entre **0** e **y**
 - la rama b aparece entre **y** y **2**
- $2 < y < +\infty$:
 - la rama a no aparece
 - la rama b no aparece

Ejemplos 17, 18

$$f_{XY}(x, y) = \begin{cases} a(x, y) & y < x < y+1, 0 < y < 2 \\ b(x, y) & y+1 < x < y+2, 0 < y < 2 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Marginación de x:

- $-\infty < x < 0$:
 - la rama a no aparece
 - la rama b no aparece
- $0 < x < 1$:
 - la rama a aparece entre **0** y **x**
 - la rama b no aparece
- $1 < x < 2$:
 - la rama a aparece entre **x-1** y **x**
 - la rama b aparece entre **0** y **x-1**
- $2 < x < 3$:
 - la rama a aparece entre **x-1** y **2**
 - la rama b aparece entre **x-2** y **x-1**
- $3 < x < 4$:
 - la rama a no aparece
 - la rama b aparece entre **x-2** y **2**
- $4 < x < +\infty$:
 - la rama a no aparece
 - la rama b no aparece

El resultado es:

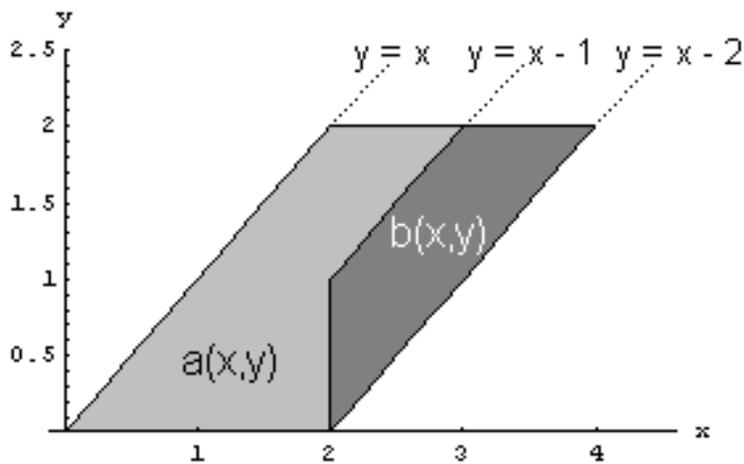
$$f_X(x) = \begin{cases} \int_0^{x-1} a(x, y) dy & 0 < x < 1 \\ \int_0^{x-1} a(x, y) dy + \int_{x-1}^2 b(x, y) dy & 1 < x < 2 \\ \int_{x-1}^2 a(x, y) dy + \int_{x-2}^0 b(x, y) dy & 2 < x < 3 \\ \int_{x-2}^2 b(x, y) dy & 3 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

Marginación de y:

- $-\infty < y < 0$:
 - la rama a no aparece
 - la rama b no aparece
- $0 < y < 2$:
 - la rama a aparece entre y e $y+1$
 - la rama b aparece entre $y+1$ e $y+2$
- $2 < y < +\infty$:
 - la rama a no aparece
 - la rama b no aparece

El resultado es:

$$f_X(x) = \begin{cases} \int_y^{y+1} a(x,y) dx + \int_{y+1}^{y+2} b(x,y) dx & 0 < y < 2 \\ 0 & \forall \text{ otro } y \end{cases}$$

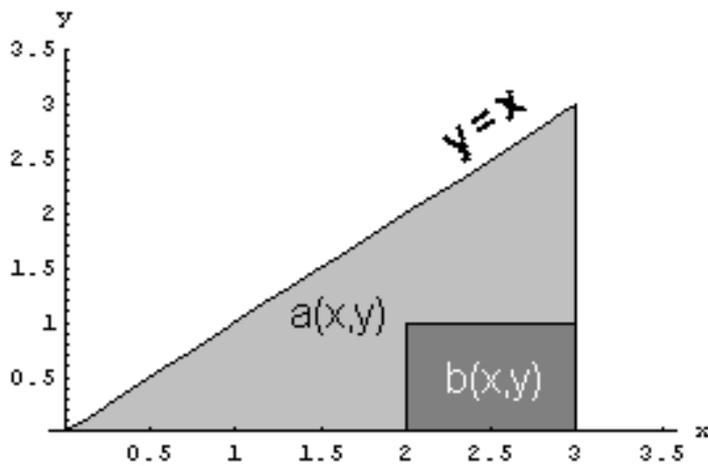


Ejemplos 19, 20

$$f_{XY}(x,y) = \begin{cases} a(x,y) & \begin{cases} 0 < x < 2, 0 < y < x \\ 2 < x < 3, x-1 < y < 2 \end{cases} \\ b(x,y) & \begin{cases} 2 < x < 3, x-2 < y < x-1 \\ 3 < x < 4, x-2 < y < 2 \end{cases} \\ 0 & \forall \text{ otro } x, y \end{cases}$$

<p>Marginación de x:</p> <ul style="list-style-type: none"> • $-\infty < x < 0$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b no aparece • $0 < x < 2$: <ul style="list-style-type: none"> • la rama a aparece entre 0 y x • la rama b no aparece • $2 < x < 3$: <ul style="list-style-type: none"> • la rama a aparece entre x-1 y 2 • la rama b aparece entre x-2 y x-1 • $3 < x < 4$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b aparece entre x-2 y 2 • $4 < x < +\infty$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b no aparece 	<p>El resultado es:</p> $f_x(x) = \begin{cases} \int_0^x a(x, y) dy & 0 < x < 2 \\ \int_{x-1}^2 a(x, y) dy + \int_{x-2}^{x-1} b(x, y) dy & 2 < x < 3 \\ \int_{x-2}^2 b(x, y) dy & 3 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$
<p>Marginación de y:</p> <ul style="list-style-type: none"> • $-\infty < y < 0$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b no aparece • $0 < y < 1$: <ul style="list-style-type: none"> • la rama a aparece entre y y 2 • la rama b aparece entre 2 e y+2 • $1 < y < 2$: <ul style="list-style-type: none"> • la rama a aparece entre y e y+1 • la rama b aparece entre y+1 e y+2 • $2 < y < +\infty$: <ul style="list-style-type: none"> • la rama a no aparece • la rama b no aparece 	<p>El resultado es:</p> $f_x(x) = \begin{cases} \int_y^2 a(x, y) dx + \int_{y+2}^2 b(x, y) dx & 0 < y < 1 \\ \int_y^{y+1} a(x, y) dx + \int_{y+1}^{y+2} b(x, y) dx & 1 < y < 2 \\ 0 & \forall \text{ otro } y \end{cases}$

Ejemplos 21, 22



$$f_{XY}(x, y) = \begin{cases} a(x, y) & \begin{cases} 0 < x < 2, 0 < y < x \\ 2 < x < 3, 1 < y < x \end{cases} \\ b(x, y) & 2 < x < 3, 0 < y < 1 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Marginación de x:

- $-\infty < x < 0$:
 - la rama a no aparece
 - la rama b no aparece
- $0 < x < 2$:
 - la rama a aparece entre **0** y **x**
 - la rama b no aparece
- $2 < x < 3$:
 - la rama a aparece entre **1** y **x**
 - la rama b aparece entre **0** y **1**
- $3 < x < +\infty$:
 - la rama a no aparece
 - la rama b no aparece

El resultado es:

$$f_X(x) = \begin{cases} \int_0^x a(x, y) dy & 0 < x < 2 \\ \int_1^x a(x, y) dy + \int_0^1 b(x, y) dy & 2 < x < 3 \\ 0 & \forall \text{ otro } x \end{cases}$$

Marginación de y:

- $-\infty < y < 0$:
 - la rama a no aparece
 - la rama b no aparece
- $0 < y < 1$:
 - la rama a aparece entre **y** y **2**
 - la rama b aparece entre **2** e **3**
- $1 < y < 3$:
 - la rama a aparece entre **y** e **3**
 - la rama b no aparece
- $3 < y < +\infty$:
 - la rama a no aparece
 - la rama b no aparece

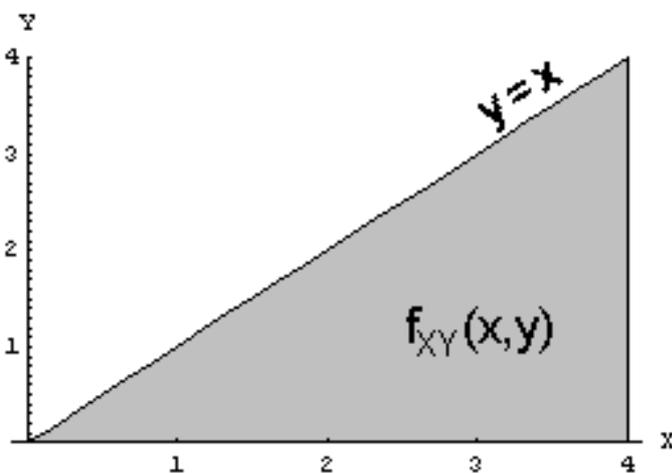
El resultado es:

$$f_x(x) = \begin{cases} \int_y^2 a(x,y)dx + \int_2^3 b(x,y)dx & 0 < y < 1 \\ \int_y^3 a(x,y)dx & 1 < y < 3 \\ 0 & \forall \text{ otro } y \end{cases}$$

Nota acerca del cálculo de probabilidades

Como hemos dicho muchas veces en esta obra, si lo que se desea es solamente calcular probabilidades, por lo general no es necesario hallar distribuciones.

Veamos un pequeño ejemplo:

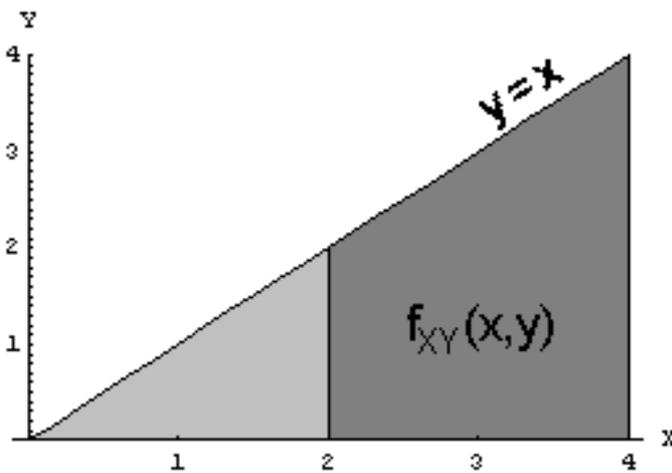


Tenemos las variables X e Y cuya distribución conjunta es:

$$f_{XY}(x,y) = \begin{cases} a(x,y) & 0 < x < 4, 0 < y < x \\ 0 & \forall \text{ otro } x, y \end{cases}$$

La probabilidad, por ejemplo, $P(Y > 2)$, es una probabilidad *marginal* (involucra solamente a la variable Y). Por lo tanto, podemos marginar para encontrar la función $f_Y(y)$ y luego obtener:

$$P(Y > 2) = \int_2^{+\infty} f_Y(y)dy$$



Pero en realidad no hace falta calcular primero la función marginal y luego integrarla, porque la probabilidad se puede hallar integrando directamente la función de densidad conjunta:

$$P(Y > 2) = \int_2^{+\infty} f_Y(y) dy$$

En el gráfico vemos que la masa de probabilidad sombreada en oscuro es la probabilidad pedida. Para más ejemplos de este tipo de cálculo, ver los problemas 6 y 7 de la sección anterior.

Problemas típicos

Los ejemplos dados en esta sección comprenden todos los problemas típicos que estudiaremos.



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/process/text/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 13 de mayo de 2004

Distribuciones condicionales

Ahora vamos a tomar lo estudiado en el primer capítulo sobre probabilidad condicional de sucesos, y lo vamos a extender a las variables aleatorias. En particular vamos a analizar cómo el hecho de que conozcamos el valor que asumió una variable aleatoria al hacer el experimento modifica la distribución de probabilidad de otra variable cuyo resultado aún no conocemos.

En el capítulo I vimos que en general, si sabemos que un **suceso** ocurre, eso modifica las probabilidades de los demás **sucesos**. Comencemos con un pequeño ejemplo:

Se realiza el experimento de tomar una persona al azar y medir su peso y su altura. Se definen los siguientes sucesos:

Suceso A: La persona pesa más de 60kg

Suceso B: La persona mide 1.90 m

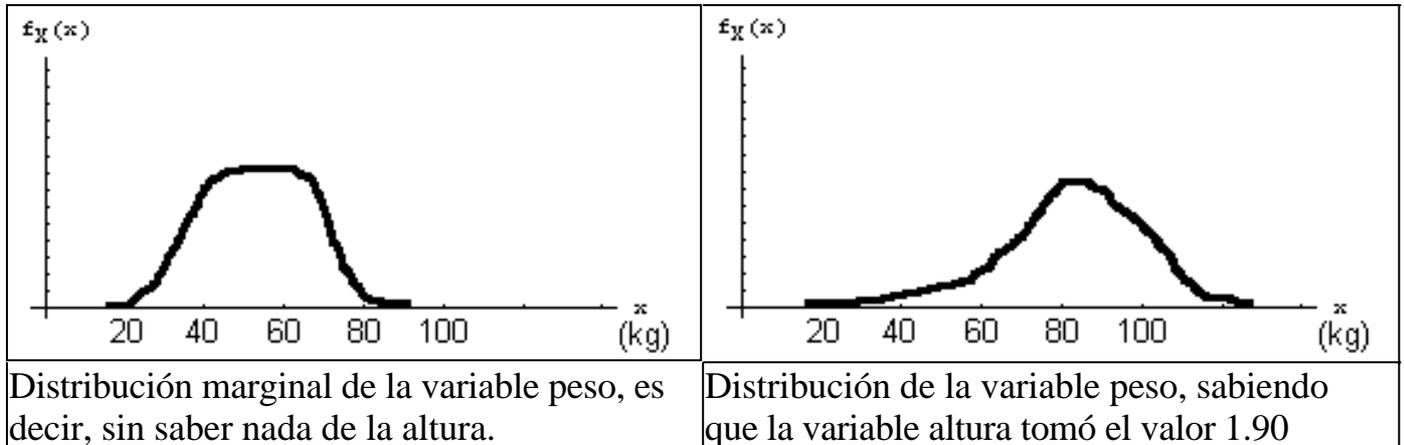
En principio el suceso A puede ocurrir con probabilidad $P(A)$. Pero si sabemos que el suceso B ocurrió, entonces la probabilidad de que ocurra A será seguramente mayor, porque si se sabe que la persona mide 1.90 m, que pese más de 60kg es más probable que si no conocemos la altura. De hecho $P(A/B)$ será un valor muy cercano a 1, porque es muy probable que una persona que *sabemos* que mide 1.90 m pese más de 60kg. Hasta aquí nada nuevo.

Ahora supongamos que el peso y la altura de la persona en realidad son **variables aleatorias**. La conclusión inmediata es que si conocemos el valor que tomó una de las variables aleatorias al hacer el experimento, eso nos modificará la distribución de probabilidad de la otra variable aleatoria.

Tenemos la función de densidad conjunta de las dos variables aleatorias. Podemos, si queremos, obtener la distribución marginal del peso, es decir, la distribución de la variable peso, que no tiene en cuenta la altura. Pero si conociéramos que la variable

altura tomó el valor 1.90m, ¿la distribución marginal del peso que teníamos sigue siendo válida?

No. Seguramente, la masa de probabilidad del peso tenderá a distribuirse más hacia los valores más altos. Un gráfico nos permitirá visualizar lo planteado y entenderlo más intuitivamente:



Podemos repetir esto muchas veces para distintos valores de la altura, y obtendríamos distintas distribuciones para el peso. Esto nos lleva a pensar que podemos encontrar una distribución "genérica" del peso en función de la altura, es decir, una función de densidad para el peso en la cual también aparezca la variable altura, y entonces para cada valor que tome la variable altura, tendremos una función de densidad distinta para el peso.

Esa distribución del peso que es genérica porque además aparece la altura, y que se transforma en una distribución en particular al darle un valor a la altura, se denomina **distribución condicional del peso dada la altura**.

En breve haremos este ejemplo con cuentas y números, pero antes enunciaremos la fórmula que hemos de utilizar:

Distribución condicional de X dado Y

Sean X, Y variables aleatorias continuas,

$$f_{X/Y}(x, y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

f_{XY} se lee "función de densidad condicional de X dado Y", y es una función de densidad de X, pero que es "genérica" porque además depende de Y, y para cada valor de Y, será una distribución en concreto para X. Es decir, **dado Y**, tenemos una distribución para X. Dicho de otro modo, **conociendo Y**, tenemos una distribución para X.

La función de densidad condicional de X dado Y determina la correspondiente distribución condicional de probabilidades, es decir, nos dice cómo se distribuyen

las probabilidades de los valores de X, una vez que se conoce el valor que ha tomado Y.

Según vemos en la fórmula, la función de densidad condicional de X se obtiene dividiendo la función de densidad conjunta por la función de densidad marginal de Y.

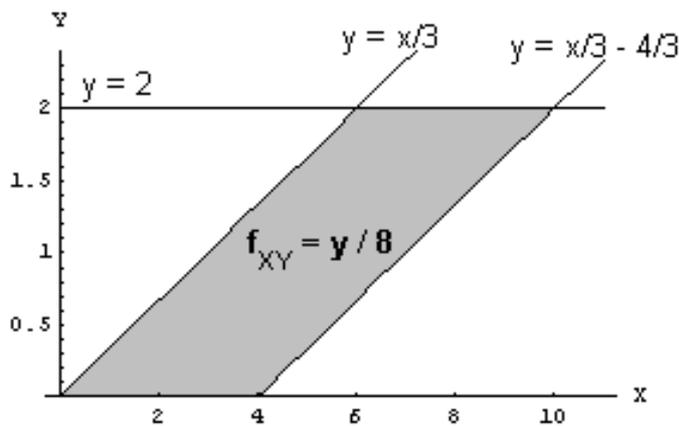
Notemos que esta fórmula es análoga a la fórmula que se dio en el capítulo I para la probabilidad condicional: $P(A/B) = P(A \cap B) / P(B)$.

Ejemplo 1

Ahora vamos a hacer el ejemplo del peso y la altura con cuentas. Por simplicidad, vamos a trabajar con distribuciones muy sencillas, que seguramente no se ajustan mucho a la realidad, pero nos permitirán visualizar los conceptos.

Llamemos X a la variable aleatoria peso, e Y a la variable aleatoria altura.

La distribución conjunta es:



$$f_{XY}(x, y) = \begin{cases} \frac{y}{8} & 0 < y < 2, \quad 3y < x < 3y + 4 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

donde Y está expresada en metros y X está expresada en decenas de kg.

Antes de hacer cuentas, observemos que las variables se condicionan mutuamente. Por ejemplo, si la altura es 2 m, el peso necesariamente está entre 60 y 100 kg (la probabilidad es nula fuera de ese intervalo). Si el peso es de 80 kg, la altura necesariamente debe estar entre 1,33 y 2m. Vemos que en principio, saber qué resultado arrojó una variable nos condiciona acerca de cuáles son los valores posibles de la otra variable. Por ejemplo en el gráfico vemos que la variable peso puede ir entre 0 y 100 kg (esto es, cuando no conocemos la altura). Pero si **conocemos** que por ejemplo la altura es 2 m, el peso ya no puede variar entre 0 y 100 kg, sino entre 60 y 100 kg.

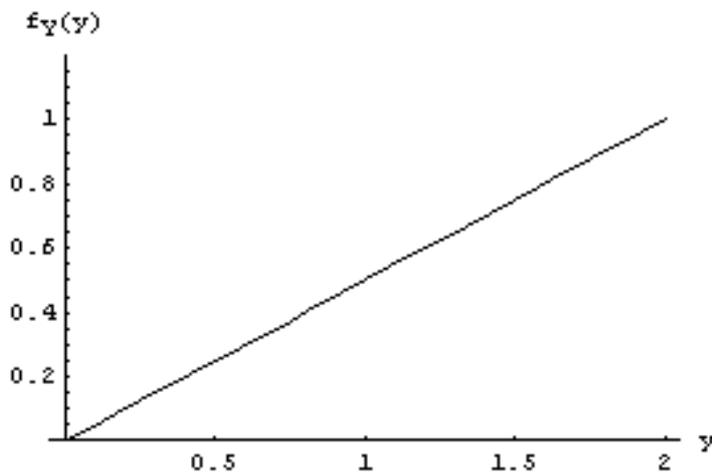
Como en este ejemplo conocemos la distribución conjunta, podríamos, por ejemplo, proceder como estudiamos en la sección anterior, para encontrar las

distribuciones marginales del peso y la altura. De esa forma tendríamos f_x y f_y , las distribuciones marginales de X e Y, es decir, las distribuciones de X y de Y que no tienen en cuenta a la otra variable. O sea, las distribuciones que tenemos para X e Y cuando no sabemos qué valor tomó la otra variable. Pero en este caso nos interesa estudiar cómo se distribuye X (el peso) si conocemos, es decir, si es dato, el valor de Y (altura).

Vamos a usar la fórmula que vimos antes:

$$f_{X/Y}(x, y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

Para encontrar la distribución condicional de X dado Y, vamos a necesitar la conjunta de X e Y, y la marginal de Y. La conjunta de X e Y es dato; la marginal de Y la encontramos a partir de la conjunta según se estudió en la sección anterior:



$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \int_{3y}^{3y+4} \frac{y}{8} dx = \frac{y}{2}$$

lo cual vale para el intervalo $0 < y < 2$. Luego:

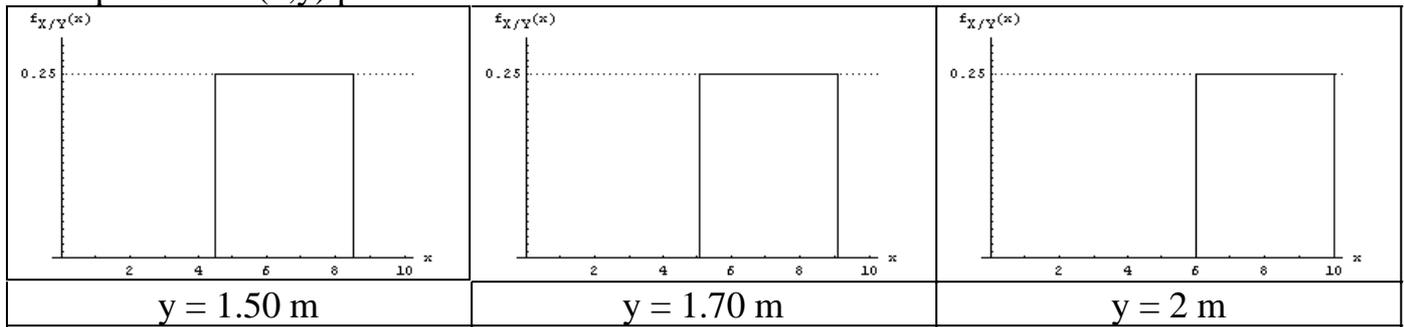
$$f_Y(y) = \begin{cases} \frac{y}{2} & 0 < y < 2 \\ 0 & \forall \text{ otro } y \end{cases}$$

Ahora vamos a dividir la conjunta por la marginal de Y para encontrar la condicional de X dado Y. Recordemos que para dividir dos funciones partidas lo que se hace es, rama a rama, dividir los valores, e intersectar los dominios. En este caso, $(y/8) / (y/2) = 1/4$, y el dominio donde esto es válido es la intersección de los dominios. Dicha intersección coincide con el dominio de la conjunta, y en dicho dominio x varía entre $3y$ y $3y+4$. En resumen queda:

$$f_{X/Y}(x, y) = \begin{cases} \frac{1}{4} & 0 < y < 2, \quad 3y < x < 3y+4 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

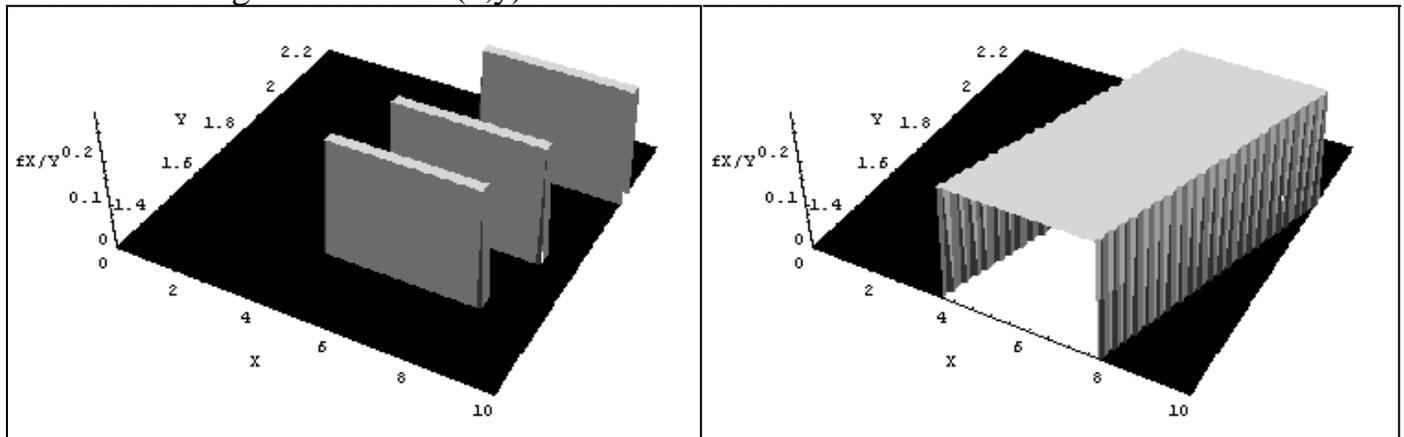
Esa es la función de densidad condicional de X dado Y. En ella podemos poner cualquier valor permitido de Y, y obtendremos la distribución de probabilidades

para X dado que conocemos el valor de Y. Por ejemplo, si en esa función ponemos $y = 1.8$, obtendremos la distribución del peso X de las personas que miden 1.80m. Grafiquemos $f_{X|Y}(x,y)$ para distintos valores de Y:



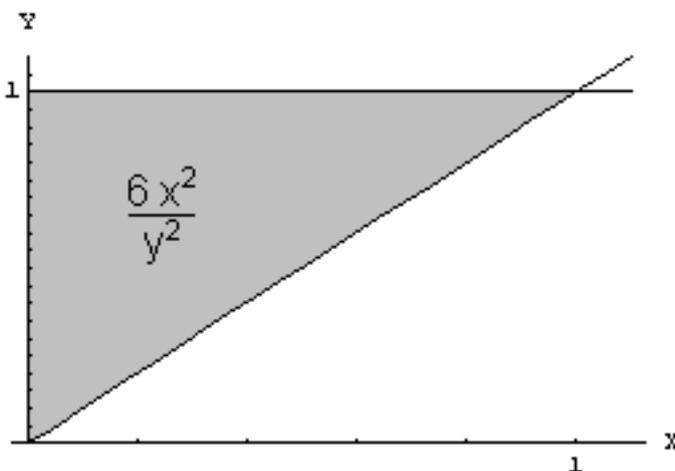
Observamos que para distintos valores de la altura, las probabilidades de los valores posibles del peso son distintas. En este caso vemos que a medida que la altura aumenta, la masa de probabilidades de los pesos se va corriendo hacia los valores grandes.

Veamos ahora gráficos de $f_{X|Y}(x,y)$ en 3 dimensiones:



Estos son los mismos cortes de antes ($y = 1.50, y = 1.70, y = 2.0$) pero vistos en 3 dimensiones

Esta es la gráfica completa, sin hacer cortes. En ella se aprecia plenamente lo que el conocimiento de la Y le hace a la X.



Ejemplo 2

$$f_{XY}(x, y) = \begin{cases} \frac{6x^2}{y^2} & 0 < y < 1, \quad 0 < x < y \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Nuevamente la consigna es hallar la distribución de X dado Y. Como en el ejemplo anterior, la distribución conjunta es dato, y debemos comenzar por hallar la marginal de Y:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \int_0^y \frac{6x^2}{y^2} dx = 2y$$

eso es válido para $0 < y < 1$, con lo cual la función de densidad marginal de y es:

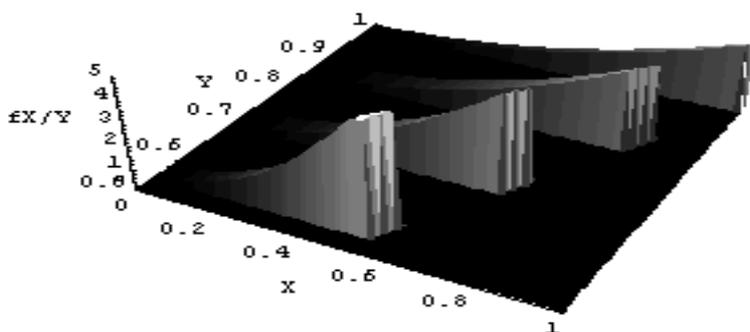
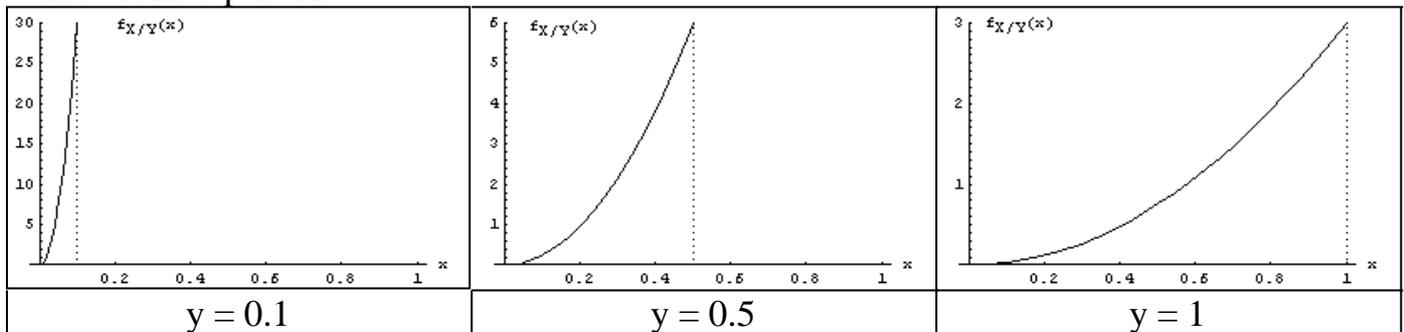
$$f_Y(y) = \begin{cases} 2y & 0 < y < 1 \\ 0 & \forall \text{ otro } y \end{cases}$$

Ahora dividimos la conjunta por la marginal de Y. Queda:

$$f_{X/Y}(x, y) = \begin{cases} \frac{3x^2}{y^3} & 0 < y < 1, \quad 0 < x < y \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Vemos que en la condicional que obtuvimos en este ejemplo, a diferencia de la anterior, la Y no aparece solamente en el dominio de la familia de funciones f_{XY} sino también en los valores.

A continuación graficaremos $f_{X/Y}$ para algunos valores de Y para visualizar nuevamente cómo el hecho de conocer el valor de Y afecta a la distribución que consideramos para X.



Por último, observamos también el efecto en un gráfico tridimensional con los cortes para $y = 0.55, 0.7, 0.85, 1$.

Cálculo de probabilidades

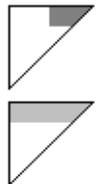
Como siempre aclaramos, para calcular probabilidades en general no es necesario hallar distribuciones. En este caso, diremos que para calcular probabilidades condicionales de variables aleatorias no hace falta encontrar las correspondientes distribuciones condicionales.

Cuando la condición es una inecuación

Supongamos que con los datos del ejemplo 2 nos piden calcular:

$$P\left(\frac{X > \frac{1}{2}}{Y > \frac{3}{4}}\right)$$

La condición es una inecuación. Esto se resuelve fácilmente mediante probabilidad condicional y calculando como se estudió al principio de este capítulo:

$$X > \frac{1}{2} \cap Y > \frac{3}{4}$$


$$P\left(\frac{X > \frac{1}{2}}{Y > \frac{3}{4}}\right) = \frac{P\left(X > \frac{1}{2} \cap Y > \frac{3}{4}\right)}{P\left(Y > \frac{3}{4}\right)} = \frac{\int_{3/4}^{1/2} \int_{3/4}^{1/2} f_{XY}(x, y) dx dy}{\int_{3/4}^1 \int_0^1 f_{XY}(x, y) dx dy}$$

Cuando la condición es una igualdad

Supongamos que con los datos del ejemplo 2 nos piden:

$$P\left(\frac{X < 0.5}{Y = 0.8}\right)$$

La condición es una igualdad. En este caso tampoco es necesario hallar la distribución condicional para calcular la probabilidad, pero veremos cómo usarla si la tenemos.

Estudiaremos 3 formas de llegar al resultado:

1) Hallando la distribución condicional y usándola:

$$P(X < 0.5 / Y = 0.8) = \int_0^{0.5} f_{X/Y=0.8}(x) dx$$

Habíamos hallado la distribución condicional en la resolución del ejemplo:

$$f_{X/Y}(x, y) = \begin{cases} \frac{3x^2}{y^3} & 0 < y < 1, \quad 0 < x < y \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Luego $f_{X/Y=0.8}(x)$ se obtiene evaluando la condicional genérica en $Y=0.8$. Obtenemos:

$$f_{X/Y=0.8}(x) = \begin{cases} \frac{3x^2}{0.512} & 0 < x < 0.8 \\ 0 & \forall \text{ otro } x \end{cases}$$

Luego el resultado es:

$$\int_0^{0.5} \frac{3x^2}{0.512} dx = 0.244$$

2) Si no tenemos la distribución condicional, no hace falta calcularla. Podemos escribir:

$$P(X < 0.5 / Y = 0.8) = \int_0^{0.5} f_{X/Y=0.8}(x) dx = \int_0^{0.5} f_{X/Y}(x, 0.8) dx = \int_0^{0.5} \frac{f_{XY}(x, 0.8)}{f_Y(0.8)} dx$$

Es decir, la función de densidad condicional evaluada en 0.8 no es otra cosa que el cociente entre la función de densidad conjunta y la marginal de Y, evaluado en 0.8. Luego, como el denominador del integrando no depende de x, lo podemos sacar de la integral.

$$\frac{1}{f_Y(0.8)} \int_0^{0.5} f_{XY}(x, 0.8) dx$$

Y asumiendo que tampoco tenemos la marginal de Y, la podemos expresar también como una integral:

$$f_Y(0.8) = f_Y(y)|_{y=0.8} = \int_0^1 f_{XY}(x, y) dx \Big|_{y=0.8} = \int_0^{0.8} f_{XY}(x, 0.8) dx$$

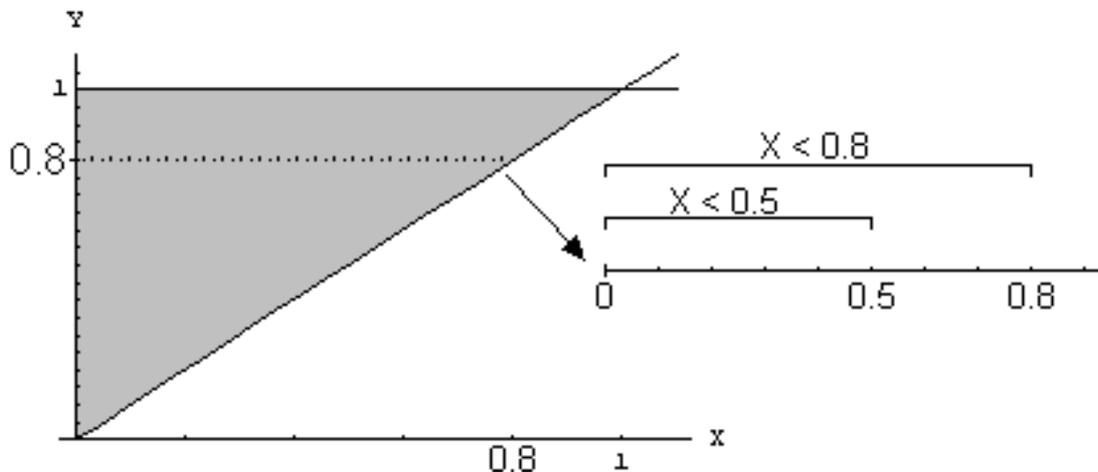
Y entonces dejamos el problema expresado como dos simples integrales de la función de densidad conjunta:

$$P(X < 0.5 / Y = 0.8) = \frac{\int_0^{0.5} f_{XY}(x, 0.8) dx}{\int_0^{0.8} f_{XY}(x, 0.8) dx} = \frac{0.3906}{1.6} = 0.244$$

Puede pensarse que esta forma de resolverlo no aporta nada, y que es lo mismo que la anterior excepto que en la anterior se hacen las integrales primero, y en esta se

hacen las integrales después. Sin embargo, esa aparente equivalencia de los problemas se debe a que las funciones usadas en el ejemplo son muy simples. Para funciones con muchas ramas y/o difíciles de integrar, la segunda forma tiene la ventaja de que hay que integrar solamente la parte del dominio que resulta útil. En cambio, para hallar las distribuciones, hay que integrar todo el dominio, y en general se usa más tiempo para dejar bien construidas las funciones.

3) Como una tercera forma, podemos apelar al hecho de que la probabilidad condicional es el cociente entre la masa de probabilidad que resulta "favorable", y la masa de probabilidad total. Pensando en eso, nos podemos "sumergir" en el universo en el cual $Y=0.8$, con lo cual se pierde una dimensión, y el dominio nos queda en un segmento de recta como vemos en el gráfico:



Luego la probabilidad de que $X < 0.5$ en ese universo, es la proporción entre la masa de probabilidad acumulada en el segmento de $0 < x < 0.5$ y la masa de probabilidad acumulada en el segmento de $0 < x < 0.8$. Es decir:

$$P\left(\frac{X < 0.5}{Y = 0.8}\right) = P_{Y=0.8}(X < 0.5) = \frac{P_{Y=0.8}(0 < X < 0.5)}{P_{Y=0.8}(0 < X < 0.8)}$$

Si nos movemos en la recta $Y = 0.8$, sabemos que $f_{xy}(x,y)$ vale siempre $f_{xy}(x,0.8)$. Luego, llegamos al mismo resultado que en la resolución anterior:

$$P(X < 0.5 / Y = 0.8) = \frac{\int_0^{0.5} f_{XY}(x, 0.8) dx}{\int_0^{0.8} f_{XY}(x, 0.8) dx}$$

Distribuciones condicionales con variables discretas

Sean X, Y variables aleatorias discretas,

$$P_{X/Y}(x, y) = \frac{P_{XY}(x, y)}{P_Y(y)}$$

P_{XY} se lee "función de probabilidad condicional de X dado Y", y es una función de probabilidad de X, pero que es "genérica" porque además depende de Y, y para cada valor de Y, será una distribución en concreto para X.

La función de probabilidad condicional de X dado Y determina la correspondiente distribución condicional de probabilidades, es decir, nos dice cómo se distribuyen las probabilidades de los valores de X, una vez que se conoce el valor que ha tomado Y.

Según vemos en la fórmula, la función de probabilidad condicional de X se obtiene dividiendo la función de probabilidad conjunta por la función de probabilidad marginal de Y. Notemos nuevamente que esta fórmula es análoga a la fórmula que se dio en el capítulo I para la probabilidad condicional: $P(A/B) = P(A \cap B) / P(B)$.

Ejemplo 3

Se tienen las variables aleatorias discretas X e Y, cuya distribución conjunta es:

X	Y		
	P_{XY}	0	2
1	0.25	0.05	0.3
2	0.15	0.1	0.15

Vamos a calcular $P_{XY}(x, y)$. Como podemos apreciar en la fórmula, vamos a necesitar la distribución marginal de Y. Le agregamos las distribuciones marginales a la tabla:

X	Y			P_X
	P_{XY}	0	2	
1	0.25	0.05	0.3	0.6
2	0.15	0.1	0.15	0.4
P_Y	0.4	0.15	0.45	

Es decir, obtuvimos que las distribuciones marginales son:

$$P_X(x) = \begin{cases} 0.6 & x=1 \\ 0.4 & x=2 \\ 0 & \forall \text{ otro } x \end{cases} \quad P_Y(y) = \begin{cases} 0.4 & y=0 \\ 0.15 & y=2 \\ 0.45 & y=4 \\ 0 & \forall \text{ otro } y \end{cases}$$

Vamos a encontrar $P_{XY}(x,y)$ de dos maneras:

1) Usando la fórmula.

La fórmula nos dice que para cada x y cada y , la probabilidad de que $X = x$ dado que $Y = y$ se obtiene como el cociente de la conjunta evaluada en (x,y) y la marginal de Y evaluada en y . Es decir, cada probabilidad de la posición (x,y) de la tabla vale $P_{XY}(x,y)/P_Y(y)$:

		Y		
		0	2	4
X	P_{XY}			
	1	0.25/0.4	0.05/0.15	0.3/0.45
	2	0.15/0.4	0.1/0.15	0.15/0.45

Haciendo las cuentas:

		Y		
		0	2	4
X	P_{XY}			
	1	5/8	1/3	2/3
	2	3/8	2/3	1/3

Llegamos al resultado fácil y rápidamente. También podemos expresar el resultado mediante alguna notación con llaves, como por ejemplo alguna de estas dos:

$$P_{X/Y}(x, y) = \begin{cases} \left. \begin{matrix} 5/8 & x=1 \\ 3/8 & x=2 \end{matrix} \right\} y=0 \\ \left. \begin{matrix} 1/3 & x=1 \\ 2/3 & x=2 \end{matrix} \right\} y=2 \\ \left. \begin{matrix} 2/3 & x=1 \\ 1/3 & x=2 \end{matrix} \right\} y=4 \\ 0 & \forall \text{ otro } x \end{cases} \quad P_{X/Y}(x, y) = \begin{cases} 5/8 & x=1 \wedge y=0 \\ 3/8 & x=2 \wedge y=0 \\ 1/3 & x=1 \wedge y=2 \\ 2/3 & x=2 \wedge y=2 \\ 2/3 & x=1 \wedge y=4 \\ 1/3 & x=2 \wedge y=4 \\ 0 & \forall \text{ otro } x \end{cases}$$

Las tres formas dadas de expresar el resultado son igualmente válidas.

2) Pensando en el significado de las distribuciones condicionales.

Notemos que tenemos 3 valores posibles para Y . Entonces tendremos 3 distribuciones condicionales para X : una para el caso $Y = 0$, otra para el caso $Y = 2$, y la otra para el caso $Y = 4$.

Primero hallaremos la distribución de X para $Y = 0$, es decir, la función $P_{XY=0}$. Esa función dará un cierto valor para $x = 1$ (la probabilidad de que $x = 1$ dado que $y = 0$), un cierto valor para $x = 2$ (la probabilidad de que $x = 2$ dado que $y = 0$) y cero para todo otro x . Con la probabilidad condicional calculamos:

$$P\left(X=1/Y=0\right) = \frac{P(X=1 \cap Y=0)}{P(Y=0)} = \frac{0.25}{0.4} = \frac{5}{8}$$

$$P\left(X=2/Y=0\right) = \frac{P(X=2 \cap Y=0)}{P(Y=0)} = \frac{0.15}{0.4} = \frac{3}{8}$$

Ahora ya podemos escribir la distribución de X dado que Y = 1:

$$P_{X/Y=1}(x) = \begin{cases} P\left(X=1/Y=1\right) & x=1 \\ P\left(X=2/Y=1\right) & x=2 \\ 0 & \forall \text{ otro } x \end{cases} = \begin{cases} \frac{5}{8} & x=1 \\ \frac{3}{8} & x=2 \\ 0 & \forall \text{ otro } x \end{cases}$$

Esto concuerda con los resultados hallados utilizando la fórmula.

De manera análoga podemos encontrar las otras dos distribuciones condicionales para X, y luego juntar las tres distribuciones y expresarlas de alguna de las maneras indicadas anteriormente (con la tabla o con las llaves).

Cálculo de probabilidades

Como dijimos para las variables continuas, para calcular probabilidades condicionales se pueden construir las distribuciones condicionales y luego usarlas para encontrar las probabilidades, o directamente encontrar las probabilidades. Con los datos del ejemplo 3, calcularemos algunas probabilidades.

Cuando las calculamos directamente, el problema se reduce a simplemente plantear la probabilidad condicional y evaluar las probabilidades conjuntas y marginales sacándolas directamente de la función de probabilidad conjunta:

$$P\left(X=1/Y>0\right) = \frac{P(X=1 \cap Y>0)}{P(Y>0)} = \frac{0.05+0.3}{0.05+0.3+0.1+0.15} = \frac{0.35}{0.6} = \frac{7}{12}$$

$$P\left(X=1/Y=2\right) = \frac{P(X=1 \cap Y=2)}{P(Y=2)} = \frac{0.05}{0.05+0.1} = \frac{0.05}{0.15} = \frac{1}{3}$$

Si contamos con la distribución condicional de X dado Y, esta última probabilidad es directamente $P_{X/Y}(1,2) = 1/3$.

Un cálculo ligeramente más complicado (en realidad es igual al primero, pero enunciado en forma más difícil):

$$\begin{aligned} P\left(X=1/Y=2 \cup Y=4\right) &= \frac{P(X=1 \cap (Y=2 \cup Y=4))}{P(Y=2 \cup Y=4)} = \frac{P(X=1 \cap Y=2) + P(X=1 \cap Y=4)}{P(Y=2) + P(Y=4)} \\ &= \frac{0.05+0.3}{0.15+0.45} = \frac{0.35}{0.6} = \frac{7}{12} \end{aligned}$$

Distribuciones con parámetros y distribuciones condicionales

Dediquemos unos momentos a observar la siguiente distribución:

$$f_X(x) = \begin{cases} \frac{1}{a} & 0 < x < a \\ 0 & \forall \text{ otro } x \end{cases}$$

Podemos advertir que para cualquier valor positivo de "a", esa expresión determina una distribución para X. Entonces en rigor lo que tenemos no es una distribución, sino una familia de distribuciones. En la práctica decimos que es una "distribución con parámetro a".

Un parámetro de una distribución es un número que aparece en la distribución, y que al tomar cada uno de sus distintos valores permitidos determina una distribución distinta para X de entre las pertenecientes a la familia.

Asignándole valores al parámetro podemos obtener distintas distribuciones. Por ejemplo:

$$f_X(x) = \begin{cases} \frac{1}{2} & 0 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_X(x) = \begin{cases} \frac{1}{5} & 0 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases} \quad \text{etc.}$$

Dicho de otro modo, **dado** un valor del parámetro, obtenemos una distribución para X.

Esto nos hace pensar que en el caso general, el parámetro "a" podría ser a su vez una variable aleatoria, con su propia distribución.

Si consideramos a "a" una variable aleatoria, debemos modificar algo en lo que escribimos originalmente. Teníamos:

$$f_X(x) = \begin{cases} \frac{1}{a} & 0 < x < a \\ 0 & \forall \text{ otro } x \end{cases}$$

Pero si "a" es una variable aleatoria en vez de una constante, entonces el miembro derecho de la ecuación no es $f_X(x)$ sino $f_{X/A}(x,a)$.

Es decir, si "a" es una variable aleatoria, la expresión que aparece en la llave, si bien distribuye a la variable aleatoria X, no es una distribución marginal, porque aparece en ella otra variable aleatoria. Deberíamos escribir:

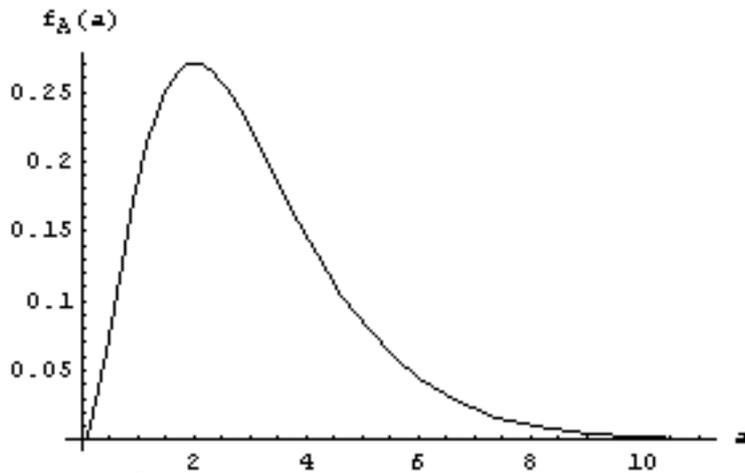
$$f_{X/A}(x) = \begin{cases} \frac{1}{a} & 0 < x < a \\ 0 & \forall \text{ otro } x \end{cases}$$

Esto nos dice que, **dado** un valor de la variable aleatoria A, tenemos una distribución para la variable aleatoria X.

En conclusión, cuándo es correcto escribir f_x y cuándo es correcto $f_{X/A}$ es una cuestión subjetiva, porque depende de si estamos considerando a A una variable aleatoria o una constante.

Ejemplo 4

Con los datos del ejemplo, supongamos además que la variable aleatoria A tiene la siguiente distribución:



$$f_A(x) = \begin{cases} \frac{1}{2} a^2 e^{-a} & a > 0 \\ 0 & a \leq 0 \end{cases}$$

Vemos que este dato es compatible con lo anterior, porque la distribución que teníamos para X requería que el parámetro "a" (que ahora lo consideramos un valor dado de A) fuera un número real positivo. Con el nuevo dato que estamos dando, todo valor posible de "a" es un número real positivo.

La función $f_{X/A}$ nos dice cómo se distribuye la variable X cuando conocemos el valor que arrojó la variable A. ¿Pero qué ocurre si no sabemos qué valor arrojó la variable A, y lo único que tenemos de ella es su distribución? ¿Podemos saber igualmente cómo se distribuye X? ¿Podemos saber cómo se distribuye X, abstrayéndonos del valor que arroje A, y teniendo en cuenta su distribución en vez de su valor arrojado que no conocemos?

Respuesta: Sí. Esa distribución que buscamos no es otra cosa que la distribución marginal de X. Eso es muy importante conceptualmente, y se recomienda que esté muy claro antes de continuar.

Procedamos entonces a calcular la distribución marginal de X. Por definición:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X/A}(x, a) da$$

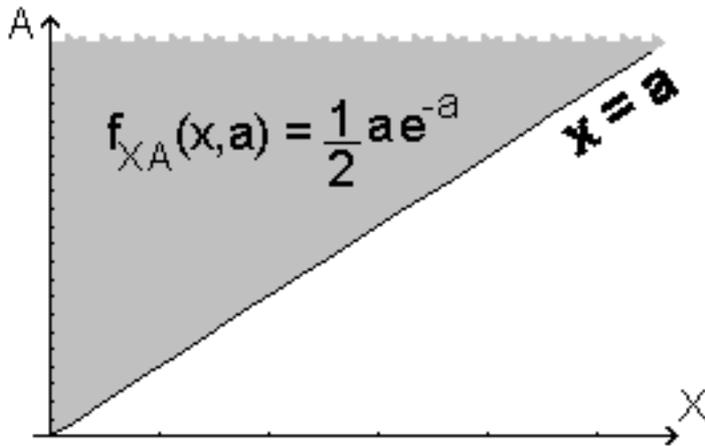
Vemos que vamos a necesitar $f_{X/A}(x, a)$, es decir la función de densidad conjunta de

X y A.

Por definición de distribución condicional:

$$f_{X/A}(x, a) = \frac{f_{XA}(x, a)}{f_A(a)}$$

Luego:



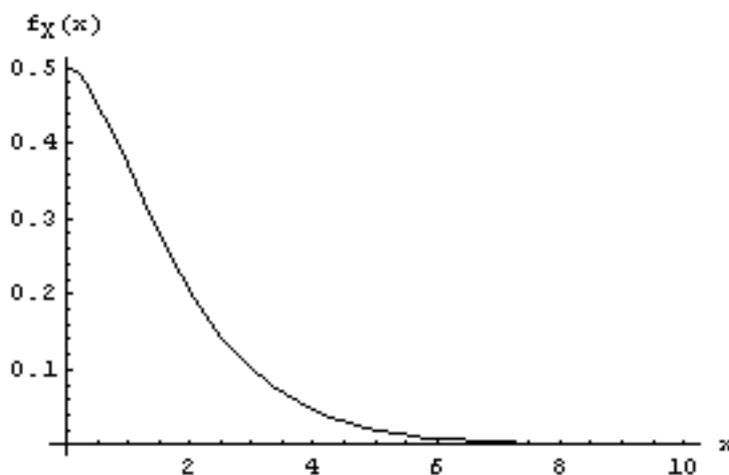
$$f_{XA}(x, a) = f_{X/A}(x, a) f_A(a) =$$

$$\begin{cases} \frac{1}{2} a e^{-a} & a > 0, 0 < x < a \\ 0 & \forall \text{ otro } x, a \end{cases}$$

Y ahora marginamos para encontrar $f_X(x)$ que es lo que buscamos:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XA}(x, a) da = \int_x^{+\infty} \frac{1}{2} a e^{-a} da = \frac{1}{2} e^{-x}(x+1)$$

Y eso vale en el intervalo $0 < x < \infty$. Luego la distribución marginal de X es:



$$f_X(x) = \begin{cases} \frac{1}{2} e^{-x}(x+1) & x > 0 \\ 0 & x \leq 0 \end{cases}$$

¿Es coherente el resultado obtenido?

- Mirando $f_{X|A}$, vemos que, dado un a , la distribución de X es no nula entre 0 y a . Además en ese intervalo es constante, es decir que la probabilidad de X no tiene a amontonarse hacia ninguna parte del intervalo $(0,a)$.
- Mirando f_A , vemos que a puede ser un valor positivo, y que la probabilidad se va haciendo más chica a medida que a se hace más grande.
- Entonces es coherente que f_X sea una función decreciente, porque X está acotada por A , y como la probabilidad de que A sea grande se va haciendo cada vez más chica, la probabilidad de que X sea grande se va haciendo cada vez más chica.

Casi todas las distribuciones que estudiaremos a partir del próximo capítulo tienen parámetros, que a veces consideraremos constantes (con lo cual la distribución de la variable será considerada marginal) y a veces las consideraremos a su vez variables aleatorias (con lo cual la distribución de la variable será condicional).

Problemas típicos

1) Se tienen las variables aleatorias discretas X e Y , cuya distribución conjunta es:

		Y			
		0	1	2	3
X	P_{XY}				
	1	0.02	0.05	0.03	0.06
	2	0.10	0.03	0.07	0.04
	4	0.02	0.08	0.04	0.15
	8	0.09	0.11	0.06	0.05

- Halle las distribuciones condicionales $P_{X|Y}$ y $P_{Y|X}$.
- Calcule $P(X = 4 / Y > 1)$
- Calcule $P(X = 2 / Y = 3)$
- Calcule $P(X > 2 / Y = 2)$
- Si se sabe que en un determinado experimento Y arrojó el valor 0, ¿cómo se distribuyen las probabilidades de X ?

Resolución

a) Procederemos igual que en el ejemplo 3. Comenzamos por encontrar las distribuciones marginales de X e Y . Las podemos anotar en los márgenes de la tabla de la conjunta, o bien aparte:

$$P_X(x) = \begin{cases} 0.16 & x = 1 \\ 0.24 & x = 2 \\ 0.29 & x = 4 \\ 0.31 & x = 8 \\ 0 & \forall \text{ otro } x \end{cases} \quad P_Y(y) = \begin{cases} 0.23 & y = 0 \\ 0.27 & y = 1 \\ 0.20 & y = 2 \\ 0.30 & y = 3 \\ 0 & \forall \text{ otro } y \end{cases}$$

Ahora encontramos las condicionales:

		Y			
		0	1	2	3
X	P _{XY}				
	1	2/23	5/27	3/20	6/30
	2	10/23	3/27	7/20	4/30
	4	2/23	8/27	4/20	15/30
	8	9/23	11/27	6/20	5/30

		Y			
		0	1	2	3
X	P _{YX}				
	1	2/16	5/16	3/16	6/16
	2	10/24	3/24	7/24	4/24
	4	2/29	8/29	4/29	15/29
	8	9/31	11/31	6/31	5/31

b) Usando probabilidad condicional y sacando los valores directamente de la función de probabilidad conjunta y de la distribución marginal de Y (o sea sin usar las condicionales):

$$P\left(X = 4 / Y > 1\right) = \frac{P(X = 4 \cap Y > 1)}{P(Y > 1)} = \frac{0.04 + 0.15}{0.5} = \frac{0.19}{0.5} = \frac{19}{50}$$

c) Usando la distribución condicional de X dado Y:

$$P\left(X = 2 / Y = 3\right) = P_{X/Y=3}(2) = P_{X/Y}(2,3) = \frac{4}{30}$$

Sin usar la distribución condicional:

$$P\left(X = 2 / Y = 3\right) = \frac{P(X = 2 \cap Y = 3)}{P(Y = 3)} = \frac{0.04}{0.30} = \frac{4}{30}$$

d) Usando la distribución condicional de X dado Y:

$$P\left(X > 2 / Y = 2\right) = P_{X/Y=2}(4) + P_{X/Y=2}(8) = P_{X/Y}(4,2) + P_{X/Y}(8,2) = \frac{4}{20} + \frac{6}{20} = \frac{1}{2}$$

Sin usar la distribución condicional:

$$P\left(X > 2 / Y = 2\right) = \frac{P(X > 2 \cap Y = 2)}{P(Y = 2)} = \frac{0.04 + 0.06}{0.2} = \frac{1}{2}$$

e) Nos están pidiendo P_X dado que Y = 0, es decir, P_{X/Y=0}.

Si ya tenemos P_{XY}, P_{X/Y=0} es una simple columna de P_{X/Y} (la columna en la cual Y=0).

Entonces copiamos los valores literalmente de P_{XY} y escribimos:

$$P_{X/Y=0}(x) = \begin{cases} 2/23 & x=1 \\ 10/23 & x=2 \\ 2/23 & x=4 \\ 9/23 & x=8 \\ 0 & \forall \text{ otro } x \end{cases}$$

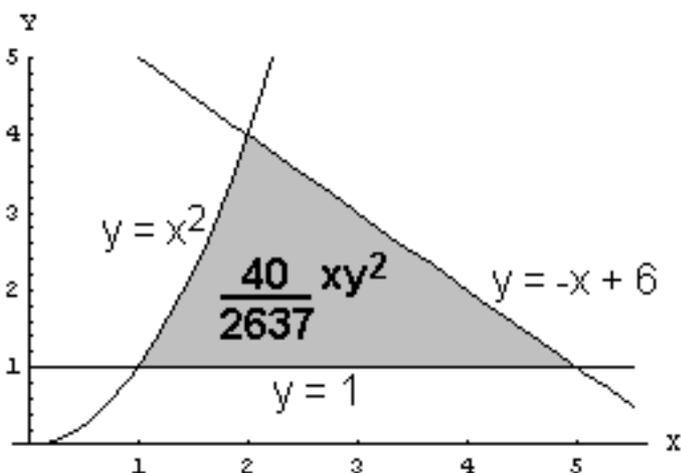
Si no tuviéramos P_{XY} , podemos usar probabilidad condicional para calcular uno por uno los cuatro valores, tomando los datos necesarios de la función de probabilidad conjunta. Por ejemplo $P_{XY=0}(1)$ se obtiene así:

$$P\left(\frac{X=1}{Y=0}\right) = \frac{P(X=1 \cap Y=0)}{P(Y=0)} = \frac{0.02}{0.23} = \frac{2}{23}$$

2) Se tienen las variables aleatorias continuas X e Y, cuya distribución conjunta es:

$$f_{XY}(x, y) = \begin{cases} \frac{40}{2637} xy^2 & y < x^2, y < -x+6, y > 1 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

- Halle las distribuciones condicionales f_{XY} y f_{YX} .
- Calcule $P(X < 3 / Y > 2)$
- Calcule $P(X > 3 / Y = 2)$
- Si se sabe que en un determinado experimento Y arrojó el valor 3, ¿cómo se distribuyen las probabilidades de X?
- Calcule $P(Y < 2 / X = 1.5)$
- Calcule $P(Y < 2 / X = 3)$
- Calcule $P(Y < 2 / X = 2)$



Resolución

- Comencemos por graficar la función de densidad conjunta para orientarnos.

Para hallar las dos distribuciones condicionales, vamos a usar las distribuciones

marginales. Por tanto, empecemos por hallarlas.

Distribución marginal de X:

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

para $1 < x < 2$: $\int_1^{x^2} \frac{40}{2637} xy^2 dy = \frac{40}{7911} x(x^6 - 1)$

para $2 < x < 5$: $\int_1^{6-x} \frac{40}{2637} xy^2 dy = \frac{40}{7911} x((6-x)^3 - 1)$

Luego:

$$f_X(x) = \begin{cases} \frac{40}{7911} x(x^6 - 1) & 1 < x < 2 \\ \frac{40}{7911} x((6-x)^3 - 1) & 2 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

Distribución marginal de Y:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \int_{\sqrt{y}}^{6-y} \frac{40}{2637} xy^2 dx = \frac{20}{2637} y^2(y^2 - 13y + 36)$$

Luego:

$$f_Y(y) = \begin{cases} \frac{20}{2637} y^2(y^2 - 13y + 36) & 1 < y < 4 \\ 0 & \forall \text{ otro } y \end{cases}$$

Ahora vamos a hallar las condicionales. Distribución condicional de X dado Y:

$$f_{X/Y}(x, y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

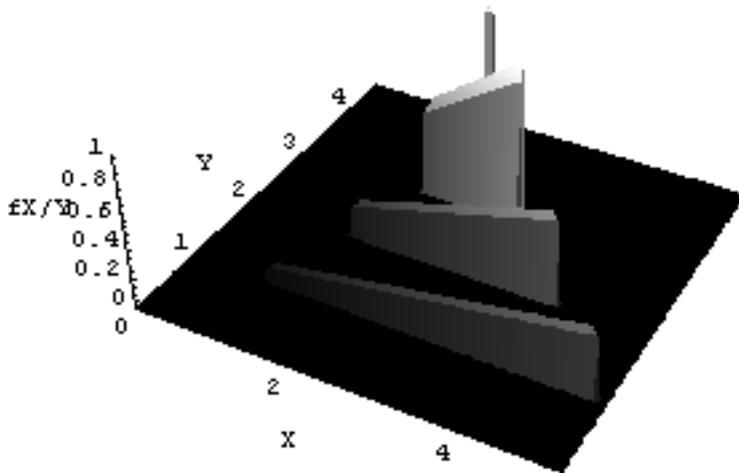
Cociente de los valores:

$$\frac{\frac{40}{2637} xy^2}{\frac{20}{2637} y^2(y^2 - 13y + 36)} = \frac{2x}{y^2 - 13y + 36}$$

Analizamos entre qué y qué varía X: $\sqrt{y} < x < 6 - y$

Luego la función de densidad condicional de X dado Y que buscábamos es:

$$f_{X/Y}(x, y) = \begin{cases} \frac{2x}{y^2 - 13y + 36} & \sqrt{y} < x < 6 - y \\ 0 & \forall \text{ otro } x, y \end{cases}$$



En el gráfico tridimensional vemos cortes de $f_{XY}(x,y)$ para $y = 1, 2, 3, 3.9$

Distribución condicional de Y dado X:

$$f_{Y/X}(x, y) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Como $f_X(x)$ es partida, el cociente entre ella y la conjunta quedará partido. Luego la condicional nos va a quedar partida.

Para $1 < x < 2$, el cociente de los valores es:

$$\frac{\frac{40}{2637}xy^2}{\frac{40}{7911}x(x^6-1)} = \frac{3y^2}{x^6-1}$$

Y tenemos que $1 < y < x^2$

Para $2 < x < 5$, el cociente de los valores es:

$$\frac{\frac{40}{2637}xy^2}{\frac{40}{7911}x((6-x)^3-1)} = \frac{3y^2}{(6-x)^3-1}$$

Y tenemos que $1 < y < 6-x$

Luego la función de densidad condicional de Y dado X que buscábamos es:

$$f_{Y/X}(x, y) = \begin{cases} \frac{3y^2}{x^6 - 1} & 1 < x < 2, 1 < y < x^2 \\ \frac{3y^2}{(6-x)^3 - 1} & 2 < x < 5, 1 < y < 6-x \\ 0 & \forall \text{ otro } x, y \end{cases}$$

b) Se obtiene aplicando probabilidad condicional e integrando directamente la función de densidad conjunta y la marginal de Y:

$$P(X < 3 / Y > 2) = \frac{P(X < 3 \cap Y > 2)}{P(Y > 2)} = \frac{\int_1^2 \int_2^{x^2} \frac{40}{2637} xy^2 dy dx + \int_2^5 \int_2^{6-x} \frac{40}{2637} xy^2 dy dx}{\int_2^5 \frac{20}{2637} y^2 (y^2 - 13y + 36) dy}$$

c) Ya que tenemos la distribución condicional de X dado Y, podemos aprovecharla:

$$P(X > 3 / Y = 2) = \int_3^{+\infty} f_{X/Y=2}(x) dx = \int_3^4 \frac{x}{7} dx = \frac{1}{2}$$

Si no tenemos la distribución condicional no hace falta que la hallemos. Podemos calcular la probabilidad pedida como lo hicimos en el ejemplo:

$$P(X > 3 / Y = 2) = \frac{\int_3^{+\infty} f_{XY}(x, 2) dx}{\int_{\sqrt{2}}^{+\infty} f_{XY}(x, 2) dx} = \frac{\int_3^4 \frac{160x}{2637} dx}{\int_{\sqrt{2}}^4 \frac{160x}{2637} dx} = \frac{1}{2}$$



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/processtext/>

d) Lo que se pide no es otra cosa que $f_{X/Y=3}(x)$. Como tenemos $f_{XY}(x, y)$, la respuesta es directamente $f_{XY}(x, 3)$, es decir:

$$f_{X/Y}(x, y) = \begin{cases} \frac{x}{3} & \sqrt{3} < x < 3 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Si no tenemos $f_{XY}(x,y)$, hacemos directamente:

$$f_{X/Y}(x,3) = \frac{f_{XY}(x,3)}{f_Y(3)}$$

y obtenemos el mismo resultado.

e) Si $X = 1.5$, vale la primera rama de $f_{Y/X}$. Luego:

$$P(Y < 2/X = 1.5) = \int_{-\infty}^2 f_{Y/X=1.5}(y) dy = \int_1^2 \frac{192}{665} y^2 dy = \frac{64}{95}$$

También se puede calcular sin tener la distribución condicional:

$$P(Y < 2/X = 1.5) = \frac{\int_1^2 f_{XY}(1.5, y) dy}{\int_1^2 f_{XY}(1.5, y) dy} = \frac{64}{95}$$

f) Si $X = 3$, vale la segunda rama de $f_{Y/X}$. Luego:

$$P(Y < 2/X = 3) = \int_{-\infty}^2 f_{Y/X=3}(y) dy = \int_1^2 \frac{3}{26} y^2 dy = \frac{7}{26}$$

También se puede calcular sin tener la distribución condicional:

$$P(Y < 2/X = 3) = \frac{\int_1^2 f_{XY}(3, y) dy}{\int_1^2 f_{XY}(3, y) dy} = \frac{7}{26}$$

g) En este caso puede haber la duda de cuál rama de $f_{Y/X}$ vale. Veamos.

Nuestro dato original era la conjunta f_{XY} . En la conjunta, la expresión que determina los valores de densidad no sufre ningún cambio en $X = 2$. Solamente el dominio cambia, y además ese cambio es continuo; es decir, no hay saltos de discontinuidad ni en el valor de la función ni en el dominio. Eso nos lleva a la conclusión de que para $X = 2$, las dos ramas de la distribución condicional tienen necesariamente que coincidir. Es como tomarle límite por izquierda y por derecha a una función continua.

Más aún, si miramos la condicional $f_{Y/X}$, y ponemos $x = 2$, veremos que las dos ramas dan el mismo valor, con lo cual se comprueba lo que suponíamos.

Entonces tenemos 3 formas de llegar al resultado. Mediante la primera rama de la condicional, mediante la segunda rama de la condicional, e integrando directamente la conjunta sin usar las condicionales. Lógicamente, de las 3 maneras se obtiene el mismo resultado.

3) Tenemos un caño de 1m de longitud. En un extremo tiene marcado el

cero, y en el otro extremo tiene marcado el uno. Se corta el tubo en una posición x al azar, y se conserva la parte que tiene el cero, descartándose el resto. Luego se vuelve a cortar el tubo en una posición y al azar, se conserva la parte que tiene el cero y se descarta el resto. Se pide:

- a) Determinar cómo se distribuye la longitud final del tubo.**
- b) Dada la longitud final del tubo, ¿cómo se distribuye la posición del primer corte?**

Resolución

X es un punto al azar entre 0 y 1, sin ninguna preferencia. Luego la distribución de X tiene que ser no nula entre 0 y 1 y además ser constante porque no hay preferencia por ningún valor. Para que sea constante entre 0 y 1 y que la integral cierre a 1, debe valer 1. Luego:

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \forall \text{ otro } x \end{cases}$$

Nos quedó un tubo que mide X . El nuevo corte es en un punto al azar entre 0 y X . Si ese punto es Y , Y debe tener probabilidad no nula entre 0 y X y ser constante (una vez fijado el X). Para que cierre a 1 la integral, la densidad debe ser $1/X$. Luego la densidad de Y es $1/X$ para $0 < y < x$, pero esa densidad es condicional porque X es una variable aleatoria. Lo que queremos decir es:

$$f_{Y/X}(x, y) = \begin{cases} \frac{1}{x} & 0 < y < x \\ 0 & \forall \text{ otro } y \end{cases}$$

En el resto del problema no hay mayores complicaciones. Nos piden f_Y . La podemos obtener marginando la conjunta, que a su vez podemos obtener multiplicando f_X y $f_{Y/X}$.

$$f_{XY}(x, y) = f_{Y/X}(x, y) f_X(x) = \begin{cases} \frac{1}{x} & 0 < y < x, 0 < x < 1 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \int_y^1 \frac{1}{x} dx = -\ln(y)$$

$$f_Y(y) = \begin{cases} -\ln(y) & 0 < y < 1 \\ 0 & \forall \text{ otro } y \end{cases}$$

Además nos piden $f_{X/Y}$. La podemos obtener dividiendo f_{XY} por f_Y .

$$f_{X/Y}(x, y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \begin{cases} \frac{-1}{x \ln(y)} & y < x < 1 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

4) Se tiran dos monedas y se llama X a la cantidad de caras que salen. Luego

Se tiran X monedas, y se llama Y a la cantidad de caras que salen. Se pide:

- Determinar la distribución de Y .
- Conociendo la cantidad de caras que salieron en la segunda tirada, ¿cómo se distribuye la cantidad de caras que salieron en la primera tirada?

Resolución

Los datos son:

$$P_X(x) = \begin{cases} \frac{1}{4} & x=0 \\ \frac{1}{2} & x=1 \\ \frac{1}{4} & x=2 \\ 0 & \forall \text{ otro } x \end{cases}$$

		Y		
		0	1	2
X	P_{YX}			
	0	1	0	0
	1	1/2	1/2	0
	2	1/4	1/2	1/4

Nos piden P_Y . La podemos obtener marginando la conjunta, que a su vez podemos obtener multiplicando P_X y P_{YX} .

Obtenemos P_{XY} de esa forma:

		Y		
		0	1	2
X	P_{XY}			
	0	1.1/4	0.1/4	0.1/4
	1	1/2.1/2	1/2.1/2	0.1/2
	2	1/4.1/4	1/2.1/4	1/4.1/4

Hacemos las cuentas:

		Y		
		0	1	2
X	P_{XY}			
	0	1/4	0	0
	1	1/4	1/4	0
	2	1/16	1/8	1/16

Marginamos y obtenemos P_Y :

$$P_Y(y) = \begin{cases} \frac{9}{16} & y=0 \\ \frac{6}{16} & y=1 \\ \frac{1}{16} & y=2 \\ 0 & \forall \text{ otro } y \end{cases}$$

Además nos piden P_{XY} . La podemos obtener dividiendo P_{XY} por P_Y .

		Y		
		0	1	2
X	P _{XY}			
	0	1/4 / 9/16	0 / 6/16	0 / 1/16
	1	1/4 / 9/16	1/4 / 6/16	0 / 1/16
	2	1/16 / 9/16	1/8 / 6/16	1/16 / 1/16

Hacemos las cuentas:

		Y		
		0	1	2
X	P _{XY}			
	0	4/9	0	0
	1	4/9	4/6	0
	2	1/9	2/6	1



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/processtext/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 13 de mayo de 2004

Independencia de variables aleatorias

En el capítulo 1 se estudió el concepto de independencia de sucesos. Se estableció que dos sucesos son estadísticamente independientes si el conocimiento de que ocurrió uno de ellos no afecta la probabilidad de que el otro ocurra.

Si quisiéramos generalizar ese concepto a las variables aleatorias, tendríamos que decir que dos variables aleatorias son estadísticamente independientes si el conocimiento del valor que arrojó una de ellas no afecta la distribución de probabilidades de los valores que puede arrojar la otra.

Pensándolo, eso es lo mismo que decir que X e Y son independientes si $f_{XY}(x,y)$ es idéntica para todos los posibles valores de y .

Yendo un paso más allá, eso es lo mismo que decir que $f_{XY}(x,y)$ no depende de y .

El siguiente paso es darse cuenta de que si $f_{XY}(x,y)$ no depende de y , entonces es en realidad $f_X(x)$, es decir, la distribución **marginal** de X , porque recordemos que hablar de distribución condicional de X tiene sentido solamente cuando en una función de densidad de X aparece alguna otra variable aleatoria.

Otra forma de pensarlo es que si $f_{XY}(x,y)$ es la distribución de X sabiendo el valor que arrojó Y , y $f_X(x)$ es la distribución de X cuando no se sabe qué valor arrojó Y , y el conocimiento de los valores de Y no afecta la distribución de X , entonces necesariamente $f_{XY}(x,y)$ debe ser igual a $f_X(x)$, porque si Y no afecta a X , entonces a la hora de distribuir X da lo mismo si conocemos el valor de Y que si no lo conocemos.

Llegamos entonces a la conclusión de que X e Y son independientes si $f_{XY}(x,y) = f_X(x)$.

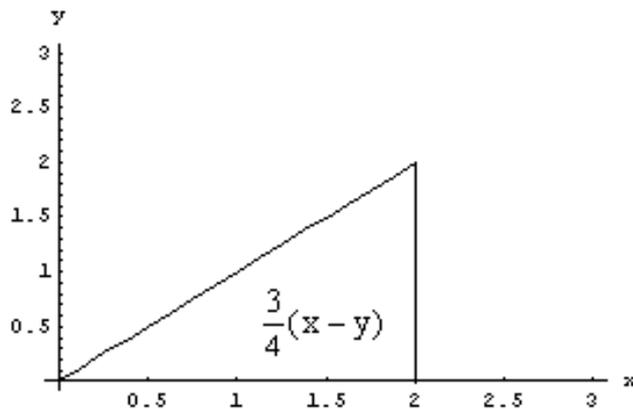
Si reemplazamos en esa ecuación $f_{XY}(x,y)$ por $f_{XY}(x,y) / f_Y(y)$, llegamos a la expresión equivalente $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

Demos entonces la definición de independencia estadística de variables aleatorias:

Para X, Y variables aleatorias continuas:	Para X, Y variables aleatorias discretas:
X e Y son estadísticamente independientes	X e Y son estadísticamente independientes
\Leftrightarrow	\Leftrightarrow
$f_{XY}(x,y) = f_X(x)$	$P_{XY}(x,y) = P_X(x)$
\Leftrightarrow	\Leftrightarrow
$f_{YX}(x,y) = f_Y(y)$	$P_{YX}(x,y) = P_Y(y)$
\Leftrightarrow	\Leftrightarrow
$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$	$P_{XY}(x,y) = P_X(x) \cdot P_Y(y)$

Independencia estadística de variables aleatorias continuas

Generalmente el dato es la distribución conjunta $f_{XY}(x,y)$. Podemos marginarla para encontrar $f_X(x)$ y $f_Y(y)$, y luego multiplicar estas últimas para ver si obtenemos de vuelta la misma $f_{XY}(x,y)$. En caso afirmativo, X e Y son independientes, y en caso negativo, no son independientes.



Ejemplo 1

$$f_{XY}(x,y) = \begin{cases} \frac{3}{4}(x-y) & 0 < x < 2, 0 < y < x \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Marginamos:

$$f_X(x,y) = \int_{-\infty}^{\infty} f_{XY}(x,y) dy = \int_0^x \frac{3}{4}(x-y) dy = \frac{3}{8}x^2$$

lo cual vale para $0 < x < 2$.

$$f_Y(x,y) = \int_{-\infty}^{\infty} f_{XY}(x,y) dx = \int_y^2 \frac{3}{4}(x-y) dx = \frac{3}{2} \left(\frac{1}{4}y^2 - y + 1 \right)$$

lo cual vale para $0 < y < 2$. Tenemos entonces:

$$f_x(x) = \begin{cases} \frac{3}{8}x^2 & 0 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_y(y) = \begin{cases} \frac{3}{2} \left(\frac{1}{4}y^2 - y + 1 \right) & 0 < y < 2 \\ 0 & \forall \text{ otro } y \end{cases}$$

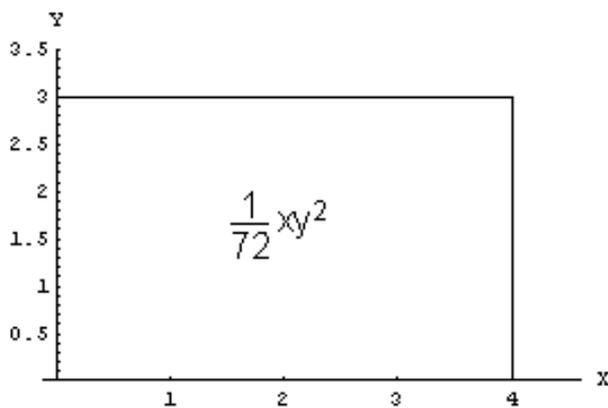
Multiplicándolas se obtiene que el valor es:

$$\frac{3}{8}x^2 \frac{3}{2} \left(\frac{1}{4}y^2 - y + 1 \right) = \frac{9}{16}x^2 \left(\frac{1}{4}y^2 - y + 1 \right)$$

Y el dominio es $0 < x < 2 \cap 0 < y < 2$.

Se ve claramente que ni los valores ni el dominio coinciden con los de la función conjunta original. Luego, **X e Y no son independientes**.

Ejemplo 2



$$f_{XY}(x, y) = \begin{cases} \frac{1}{72}xy^2 & 0 < x < 4, 0 < y < 3 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Marginamos:

$$f_x(x, y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_0^3 \frac{1}{72}xy^2 dy = \frac{1}{8}x$$

lo cual vale para $0 < x < 4$.

$$f_y(x, y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \int_0^4 \frac{1}{72}xy^2 dy = \frac{1}{9}y^2$$

lo cual vale para $0 < y < 3$. Tenemos entonces:

$$f_x(x) = \begin{cases} \frac{1}{8}x & 0 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_y(y) = \begin{cases} \frac{1}{9}y^2 & 0 < y < 3 \\ 0 & \forall \text{ otro } y \end{cases}$$

Multiplicándolas se obtiene que el valor es:

$$\frac{1}{8}x \frac{1}{9}y^2 = \frac{1}{72}xy^2$$

Y el dominio es $0 < x < 4 \cap 0 < y < 3$.

Vemos que tanto el valor de la función como el dominio coinciden con los de la

función de densidad conjunta original. Luego, **X e Y son independientes**.

Consideraciones acerca del dominio

Hay una manera que en algunos casos permite determinar en forma inmediata y sin hacer cuentas que dos variables no son independientes.

Observemos el dominio de la función del ejemplo 1. Si supiéramos que X vale 1, entonces Y puede asumir cualquier valor entre 0 y 1. Si supiéramos que X vale 2, entonces Y puede asumir cualquier valor entre 0 y 2. Vemos entonces que el hecho de conocer el valor que arrojó X nos afecta cuáles son los valores posibles de Y. Entonces es evidente que X e Y **no son independientes**.

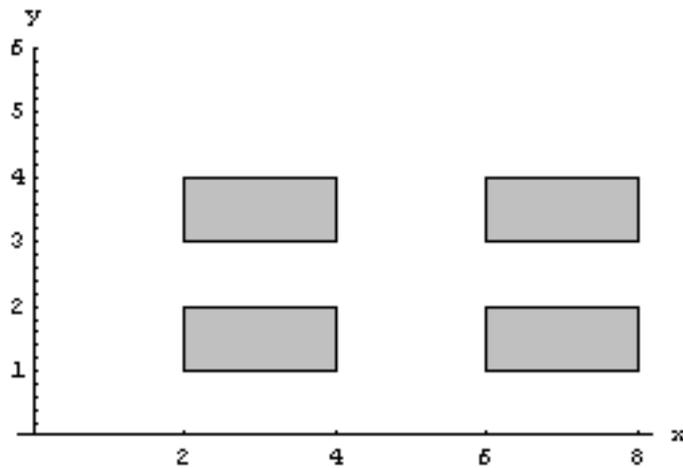
Miremos en cambio el dominio de la función del ejemplo 2. La variable Y puede asumir cualquier valor entre 0 y 3, sin importar el valor que haya arrojado X.

Análogamente, saber cuánto vale Y tampoco condiciona los valores posibles de X. ¿Esto significa que son independientes? **NO**. Solamente significa que el conjunto de valores posibles de cada variable no es afectado por el conocimiento del valor que arrojó la otra. Pero lo que sí puede cambiar es cómo se distribuye la probabilidad entre los valores posibles. Entonces puede que no sean independientes.

¿Qué característica del dominio del ejemplo 2 es la que hace que el conjunto de valores posibles de cada variable no sea afectado por el conocimiento del valor que arrojó la otra variable? Que tiene forma **rectangular**. Entonces por lo que dijimos antes, que el dominio sea rectangular es **condición necesaria** para que las variables sean independientes. Pero no suficiente.

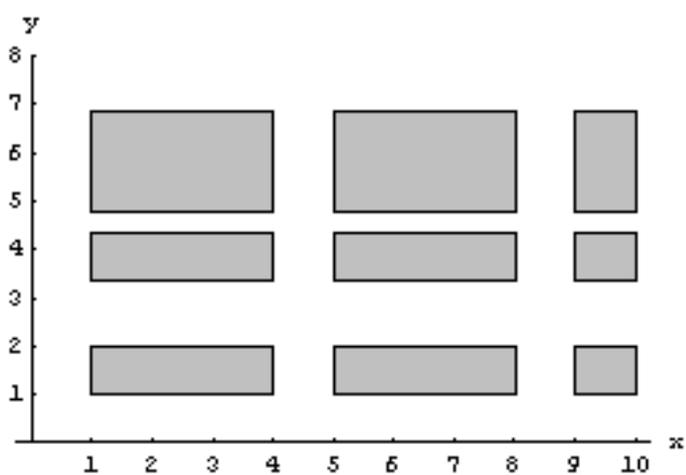
Entonces, en el ejemplo 1, con solo mirar el dominio podríamos haber contestado que las variables no son independientes, sin hacer ninguna cuenta. En el ejemplo 2, vemos que pueden ser independientes porque el dominio lo permite, pero también podrían no serlo, por lo cual hay que hacer la cuenta para determinarlo.

Dijimos que el hecho de que el dominio tenga forma rectangular es condición necesaria (pero no suficiente) para que las variables sean independientes. Ahora vamos a analizar un poco más en detalle qué significa que el dominio tenga "forma rectangular". Lo que vamos a ver es que algo puede tener "forma rectangular" sin ser un rectángulo. Consideremos un



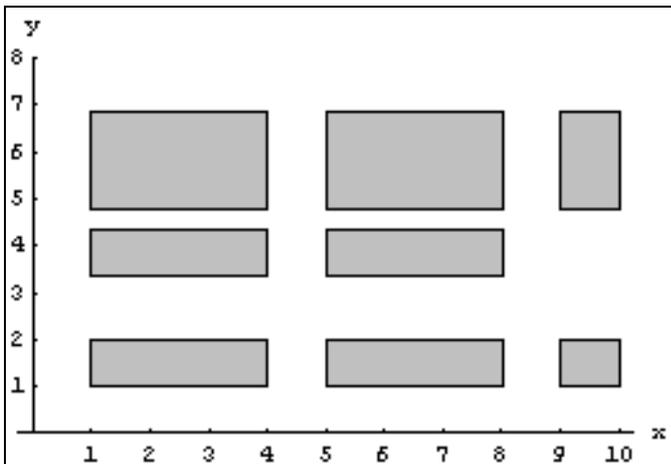
par de variables X, Y tales que el dominio de su función de densidad conjunta es como el que se ve en el gráfico. Ese dominio, ¿impide que las variables sean independientes? Si X arrojará el valor 3, Y podría estar entre 1 y 2, ó entre 3 y 4. Si X arrojará el valor 6.2, Y podría estar entre 1 y 2, ó entre 3 y 4. De hecho para cualquier valor que pueda tomar X , Y puede estar entre 1 y 2, ó entre 3 y 4. Y para cualquier valor que pueda tomar Y , X puede estar entre 2 y 4 ó entre 6 y 8. Vemos entonces que la forma de este dominio no impide que las variables sean independientes, y no es un rectángulo, sino que solamente tiene **forma rectangular**.

Veamos otro ejemplo(4) de un dominio con forma rectangular:

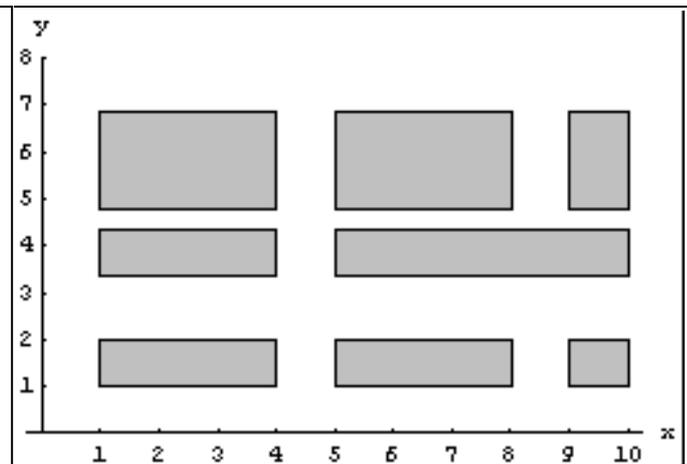


Observándolo cuidadosamente vemos que este otro dominio tampoco impide la independencia de las variables. Notemos que ni siquiera está formado por rectángulos todos iguales, pero sí se verifica que para cualquier X los valores de Y varían entre los mismos límites, y que para cualquier Y los valores de X varían entre los mismos límites.

Veamos ahora dos ejemplos de dominios que impiden la independencia aunque tienen "forma rectangular":



Ejemplo 5: Si $X = 6$, los valores posibles de Y son distintos que si $X = 9.5$. Luego X e Y no pueden ser independientes.



Ejemplo 6: Si $X = 2$, los valores posibles de Y son distintos que si $X = 8.5$. Luego X e Y no pueden ser independientes.

¿Cuál podría ser entonces la regla práctica? Simplemente observaremos el dominio y veremos si los valores de una de las variables condicionan los valores de la otra. Si los condicionan, entonces las variables obviamente no son independientes. Si no los condicionan, haremos las marginaciones y veremos si lo son o no.

Entrar en un mayor nivel de rigurosidad matemática no tiene sentido, porque en ese caso la regla dejaría de ser práctica. Solamente haremos el comentario para quien esté interesado, de que para que un dominio permita la independenciam, debe poder expresarse con una lista de condiciones que involucren a una sola variable por vez:

Ejemplo	Dominio	Independientes
1	$0 < x < 2 \cap 0 < y < x$	no pueden serlo
2	$0 < x < 4 \cap 0 < y < 3$	pueden serlo
3	$(2 < x < 4 \cup 6 < x < 8) \cap (1 < y < 2 \cup 3 < y < 4)$	pueden serlo
4	$(1 < x < 4 \cup 5 < x < 8 \cup 9 < x < 10) \cap (1 < y < 2 \cup 3.3 < y < 4.3 \cup 4.9 < y < 6.9)$	pueden serlo
5	$(1 < x < 4 \cup 5 < x < 8 \cup 9 < x < 10) \cap (1 < y < 2 \cup 3.3 < y < 4.3 \cup 4.9 < y < 6.9 \text{ si } x < 8, 1 < y < 2 \cup 4.9 < y < 6.9 \text{ si } x > 8)$	no pueden serlo
6	$(1 < y < 2 \cup 3.3 < y < 4.3 \cup 4.9 < y < 6.9) \cap (1 < x < 4 \cup 5 < x < 10 \text{ si } y < 3 \cup y > 4.3, 1 < x < 4 \cup 5 < x < 8 \cup 9 < x < 10 \text{ si } 3 < y < 4.3)$	no pueden serlo

Evidentemente, en la práctica nunca perderemos tiempo haciendo este análisis para ver si es posible que las variables lleguen a ser independientes, porque resultaría más rápido directamente hacer las cuentas de marginación y ver concretamente si lo son o no.

Independencia estadística de variables aleatorias discretas

Generalmente el dato es la distribución conjunta $P_{xy}(x,y)$. Podemos marginarla para

encontrar $P_X(x)$ y $P_Y(y)$, y luego multiplicar estas últimas para ver si obtenemos de vuelta la misma $f_{XY}(x,y)$. En caso afirmativo, X e Y son independientes, y en caso negativo, no son independientes.

Ejemplo 7

Tenemos las variables aleatorias discretas X e Y, cuya distribución conjunta es:

X	Y		
	P_{XY}	1	2
1	0.12	0.1	0.08
2	0.28	0.2	0.22

Hallamos las distribuciones marginales:

X	Y			P_X
	P_{XY}	1	2	
1	0.12	0.1	0.08	0.3
2	0.28	0.2	0.22	0.7
P_Y	0.4	0.3	0.3	---

Si multiplicamos las distribuciones marginales obtenemos:

X	Y		
	$P_X P_Y$	1	2
1	0.12	0.09	0.09
2	0.28	0.21	0.21

Vemos que $P_X P_Y \neq P_{XY}$, por lo tanto X e Y **no son independientes**.

Aunque que para algunos valores se cumple (por ejemplo, para (1,1) y para (2,1)) no se cumple para todos. Con que no se cumpla para un solo valor, ya no son independientes.

Ejemplo 8

Tenemos las variables aleatorias discretas X e Y, cuya distribución conjunta es:

X	Y		
	P_{XY}	1	2
1	0.08	0.12	0.2
2	0.12	0.18	0.3

Hallamos las distribuciones marginales:

X	Y			P_X
	P_{XY}	1	2	
1	0.08	0.12	0.2	0.4
2	0.12	0.18	0.3	0.6
P_Y	0.2	0.3	0.5	---

Si multiplicamos las distribuciones marginales obtenemos:

X	P _X P _Y	Y		
		1	2	3
1	0.08	0.12	0.2	
2	0.12	0.18	0.3	

Vemos que $P_X P_Y = P_{XY} \forall x, y$. Por lo tanto X e Y **son independientes**.

Problemas típicos

Como problemas típicos de independencia de dos variables discretas, se pueden tomar los ejemplos 7 y 8. Los siguientes problemas son con variables continuas.

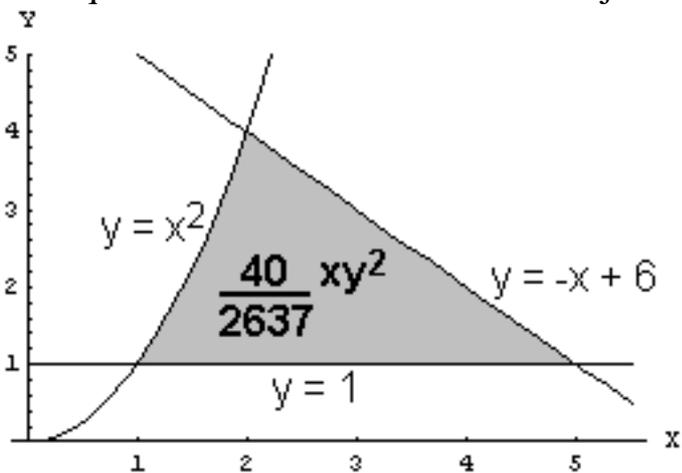
1) Se tienen las variables aleatorias continuas X e Y, cuya distribución conjunta es:

$$f_{XY}(x, y) = \begin{cases} \frac{40}{2637} xy^2 & y < x^2, y < -x + 6, y > 1 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Determine si son independientes.

Resolución

Grafiquemos la función de densidad conjunta:



Vemos que el dominio no tiene forma rectangular. Por lo tanto, los valores de una de las variables afectan los valores de la otra. Luego, **no son independientes**.

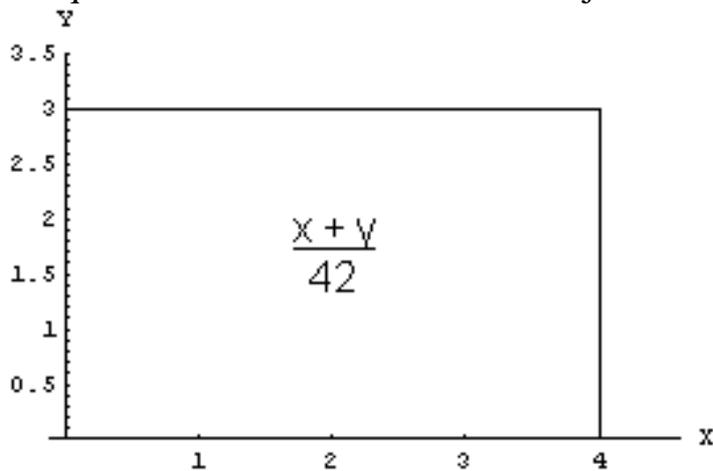
2) Se tienen las variables aleatorias continuas X e Y, cuya distribución conjunta es:

$$f_{XY}(x, y) = \begin{cases} \frac{x+y}{42} & 0 < x < 4, 0 < y < 3 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Determine si son independientes.

Resolución

Grafiquemos la función de densidad conjunta:



Vemos que el dominio es rectangular. Por lo tanto, es posible que las variables sean independientes. Hagamos las cuentas para determinarlas:

$$f_X(x, y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_0^3 \frac{x+y}{42} dy = \frac{3+2x}{28}$$

lo cual vale para $0 < x < 4$.

$$f_Y(x, y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \int_0^4 \frac{x+y}{42} dx = \frac{4+2y}{21}$$

lo cual vale para $0 < y < 3$. Tenemos entonces:

$$f_X(x) = \begin{cases} \frac{3+2x}{28} & 0 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_Y(y) = \begin{cases} \frac{4+2y}{21} & 0 < y < 3 \\ 0 & \forall \text{ otro } y \end{cases}$$

Multiplicándolas se obtiene:

$$\frac{3+2x}{28} \frac{4+2y}{21} \neq \frac{x+y}{42}$$

Es decir, $f_X f_Y \neq f_{XY}$. Luego, X e Y **no son independientes**.

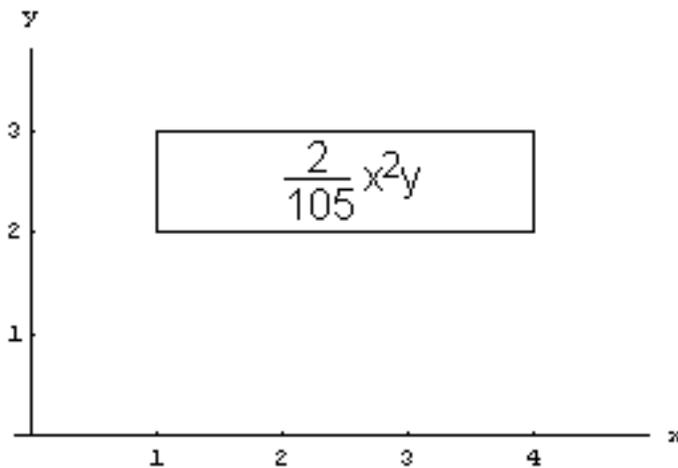
3) Se tienen las variables aleatorias continuas X e Y, cuya distribución conjunta es:

$$f_{XY}(x, y) = \begin{cases} \frac{2}{105} x^2 y & 1 < x < 4, 2 < y < 3 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Determine si son independientes.

Resolución

Grafiquemos la función de densidad conjunta:



Vemos que el dominio es rectangular. Por lo tanto, es posible que las variables sean independientes. Hagamos las cuentas para determinarlo:

$$f_X(x, y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_2^3 \frac{2}{105} x^2 y dy = \frac{1}{21} x^2$$

lo cual vale para $1 < x < 4$.

$$f_Y(x, y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \int_1^4 \frac{2}{105} x^2 y dx = \frac{2}{5} y$$

lo cual vale para $2 < y < 3$. Tenemos entonces:

$$f_X(x) = \begin{cases} \frac{1}{21} x^2 & 1 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_Y(y) = \begin{cases} \frac{2}{5} y & 2 < y < 3 \\ 0 & \forall \text{ otro } y \end{cases}$$

Multiplicándolas se obtiene que el valor es:

$$\frac{1}{21} x^2 \frac{2}{5} y = \frac{2}{105} x^2 y$$

Y el dominio es $0 < x < 4 \cap 0 < y < 3$.

Vemos que tanto el valor de la función como el dominio coinciden con los de la función de densidad conjunta original. Luego, **X e Y son independientes**.

4) El capataz de una obra le asigna una tarea a un obrero y otra tarea a otro obrero. El tiempo en horas que tarda el obrero 1 en completar la tarea que le asignaron es la variable aleatoria X, y el tiempo en horas que tarda el obrero 2 en completar la tarea que le asignaron es la variable aleatoria Y.

Las distribuciones de X e Y son:

$$f_X(x) = \begin{cases} \frac{1}{3} & 0 < x < 3 \\ 0 & \forall \text{ otro } x \end{cases} ; \quad f_Y(y) = \begin{cases} \frac{1}{8} y & 0 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

¿Cuál es la probabilidad de que ambos tarden más de una hora en completar sus respectivas tareas? ¿Qué condición debe cumplirse para que el problema se pueda resolver?

Resolución

Nos están pidiendo $P(X > 1 \cap Y > 1)$. La condición que debe cumplirse para que el problema se pueda resolver es que las variables X e Y sean independientes. Hay dos formas para calcular la probabilidad pedida, y ambas requieren esa condición.

Una forma posible es, asumiendo independencia, hacer:

$$P(X > 1 \cap Y > 1) = P(X > 1) P(Y > 1)$$

Luego el resultado es el producto de dos integrales:

$$P(X > 1 \cap Y > 1) = \int_1^{+\infty} f_X(x) dx \cdot \int_1^{+\infty} f_Y(y) dy = \int_1^3 \frac{1}{3} dx \cdot \int_1^4 \frac{1}{8} y dy = \frac{2}{3} \cdot \frac{15}{16} = \frac{5}{8}$$

Otra forma posible es darse cuenta de que:

$$P(X > 1 \cap Y > 1) = \int_1^{+\infty} \int_1^{+\infty} f_{XY}(x, y) dy dx$$

Pero para resolverlo vamos a necesitar f_{XY} , y no la tenemos. La única forma de obtener f_{XY} si solamente contamos con f_X y f_Y , es asumir que X e Y son independientes, y de esa forma se obtiene $f_{XY} = f_X \cdot f_Y$.

Asumiendo independencia, obtenemos:

$$f_{XY}(x, y) = \begin{cases} \frac{1}{24} y & 0 < x < 3, 0 < y < 4 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Luego:

$$P(X > 1 \cap Y > 1) = \int_1^3 \int_1^4 \frac{1}{24} y dy dx = \frac{5}{8}$$



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/process/text/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 13 de mayo de 2004

Esperanza condicional y Regresión

Para definir la esperanza condicional, vamos a combinar dos conceptos que ya hemos estudiado: la esperanza de una distribución, y la distribución condicional.

Dada una distribución, su media o esperanza nos da una idea de cuál es el valor que podemos esperar obtener al hacer el experimento. A su vez, la distribución condicional es un modelo que, dado el valor arrojado por una variable, nos permite tener una distribución de probabilidades para la otra variable.

La función de densidad condicional, por ejemplo de X dado Y, depende de "x" y de "y", y nos permite obtener una **distribución para X**, al conocer el valor de Y. Podemos pensar que el "y" que aparece en $f_{X|Y}(x,y)$ es simplemente un número, un parámetro, ya que para cualquier valor válido de "y", $f_{X|Y}(x,y)$ es una distribución perfectamente válida para x. Recordemos que la diferencia entre $f_{X|A}(x,a)$ y una $f_X(x,a)$ es si consideramos a A una variable aleatoria o simplemente un parámetro.

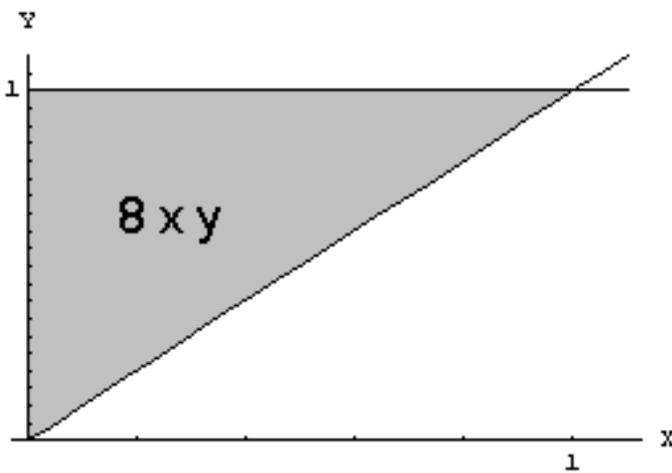
Por lo tanto, podemos calcularle la esperanza a $f_{X|Y}(x,y)$ asumiendo que "y" es simplemente un número. Luego, obtendremos una esperanza para X que dependerá de "y". Esta herramienta sirve ver cómo los valores de Y afectan al valor esperado de X.

Así como la esperanza de la distribución $f_X(x)$ es:

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

análogamente la esperanza de la distribución $f_{X|Y}(x)$ es:

$$E(X/Y) = \mu_{X/Y} = \int_{-\infty}^{\infty} x f_{X/Y}(x, y) dx$$



Ejemplo

Se tienen las variables aleatorias X e Y, cuya distribución conjunta es:

$$f_{XY}(x, y) = \begin{cases} 8xy & 0 < x < 1, \quad x < y < 1 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Podemos obtener las distribuciones marginales:

$$f_X(x) = \begin{cases} 4x(1-x^2) & 0 < x < 1 \\ 0 & \forall \text{ otro } x \end{cases}; \quad f_Y(y) = \begin{cases} 4y^3 & 0 < y < 1 \\ 0 & \forall \text{ otro } y \end{cases}$$

Y las condicionales:

$$f_{X/Y}(x, y) = \begin{cases} \frac{2x}{y^2} & 0 < x < y, \quad 0 < y < 1 \\ 0 & \forall \text{ otro } x, y \end{cases};$$

$$f_{Y/X}(x, y) = \begin{cases} \frac{2y}{1-x^2} & x < y < 1, \quad 0 < x < 1 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Calculemos la esperanza condicional de X dado Y:

$$\mu_{X/Y} = \int_{-\infty}^{\infty} x f_{X/Y}(x, y) dx = \int_0^y \frac{2x^2}{y^2} dx = \frac{2}{3}y$$

Eso quiere decir, que si por ejemplo la variable Y arroja el valor 1/2, el valor esperado para X será 1/3. Es decir, si conocemos el valor que arrojó Y, el valor esperado de X es 2/3 de ese valor.

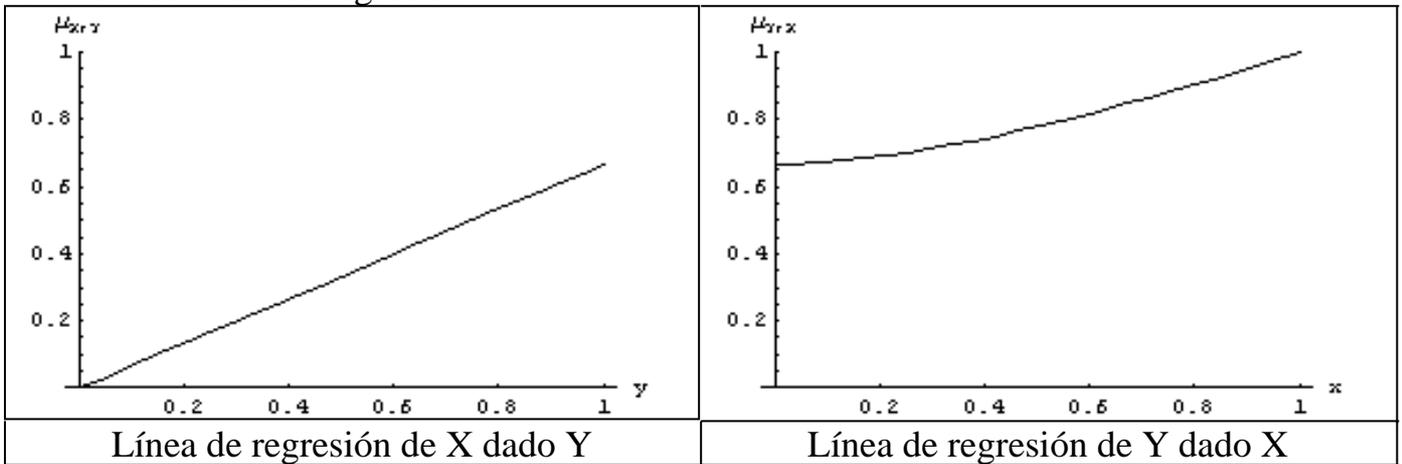
También podemos calcular la esperanza condicional de Y dado X:

$$\mu_{Y/X} = \int_{-\infty}^{\infty} y f_{Y/X}(x, y) dx = \int_x \frac{2y^2}{1-x^2} dy = \frac{2(x^2 + x + 1)}{3(x+1)}$$

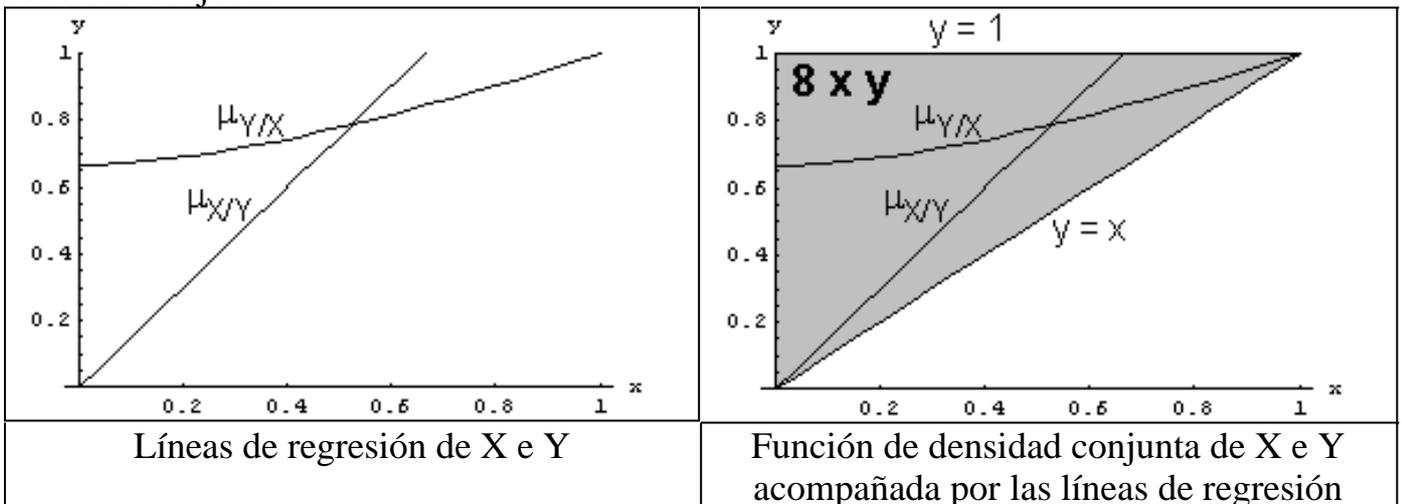
Líneas de regresión

Como se estudió en secciones anteriores, la esperanza de una variable, por ejemplo μ_x , es un número, es decir una constante. Pero la esperanza condicional es una función, no una constante. Por ejemplo, la esperanza condicional de X dado Y, $\mu_{X/Y}$, es una función de "y". Luego, nos puede interesar hacer su gráfico, con "y" en un eje, y $\mu_{X/Y}$ en el otro. La línea que resulta de graficar $\mu_{X/Y}(y)$ se denomina **línea de regresión de X dado Y**.

Análogamente, la gráfica de $\mu_{Y/X}(x)$ se denomina línea de regresión de Y dado X. Veamos las líneas de regresión:



También es habitual dibujarlas juntas, e incluso dibujarlas sobre la función de densidad conjunta:



¿Qué sucede cuando las variables son independientes?

Si X e Y son independientes, $f_{XY}(x,y)$ es directamente $f_X(x)$. Consecuentemente, la

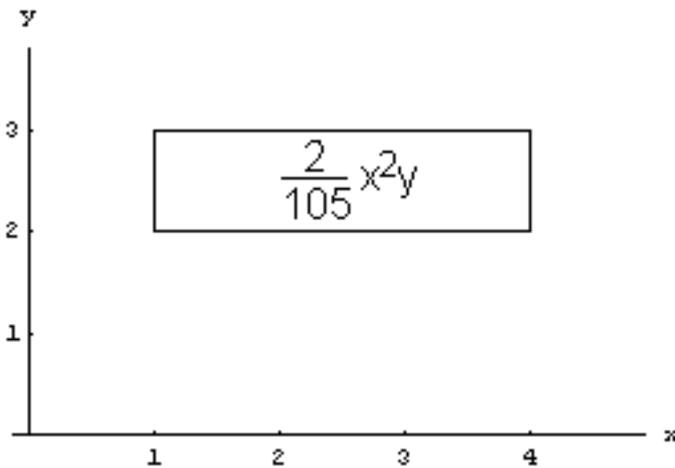
esperanza condicional de X dado Y, $\int_{-\infty}^{\infty} x f_{X/Y}(x, y) dx$, es directamente

$$\int_{-\infty}^{\infty} x f_X(x) dx$$

, es decir, la esperanza de X. Esto es razonable, ya que si las variables son independientes, el valor que arroja Y no afecta el valor esperado para X.

Las líneas de regresión serán entonces rectas y sin pendiente, porque serán la gráfica de una constante. Por ejemplo la línea de regresión de X dado Y, es decir la gráfica de $\mu_{X/Y}$, al ser las variables independientes resulta ser directamente la gráfica de μ_X , que no es una función sino un número.

Ejemplo



Se tienen las variables aleatorias X e Y, cuya distribución conjunta es:

$$f_{XY}(x, y) = \begin{cases} \frac{2}{105} x^2 y & 1 < x < 4, 2 < y < 3 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Hallamos las distribuciones marginales:

$$f_X(x) = \begin{cases} \frac{1}{21} x^2 & 1 < x < 4 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_Y(y) = \begin{cases} \frac{2}{5} y & 2 < y < 3 \\ 0 & \forall \text{ otro } y \end{cases}$$

Vemos que si multiplicamos f_X y f_Y obtenemos f_{XY} . Luego las variables son independientes, y entonces, como ya se demostró, f_{XY} y $f_{Y/X}$ son directamente f_X y f_Y .

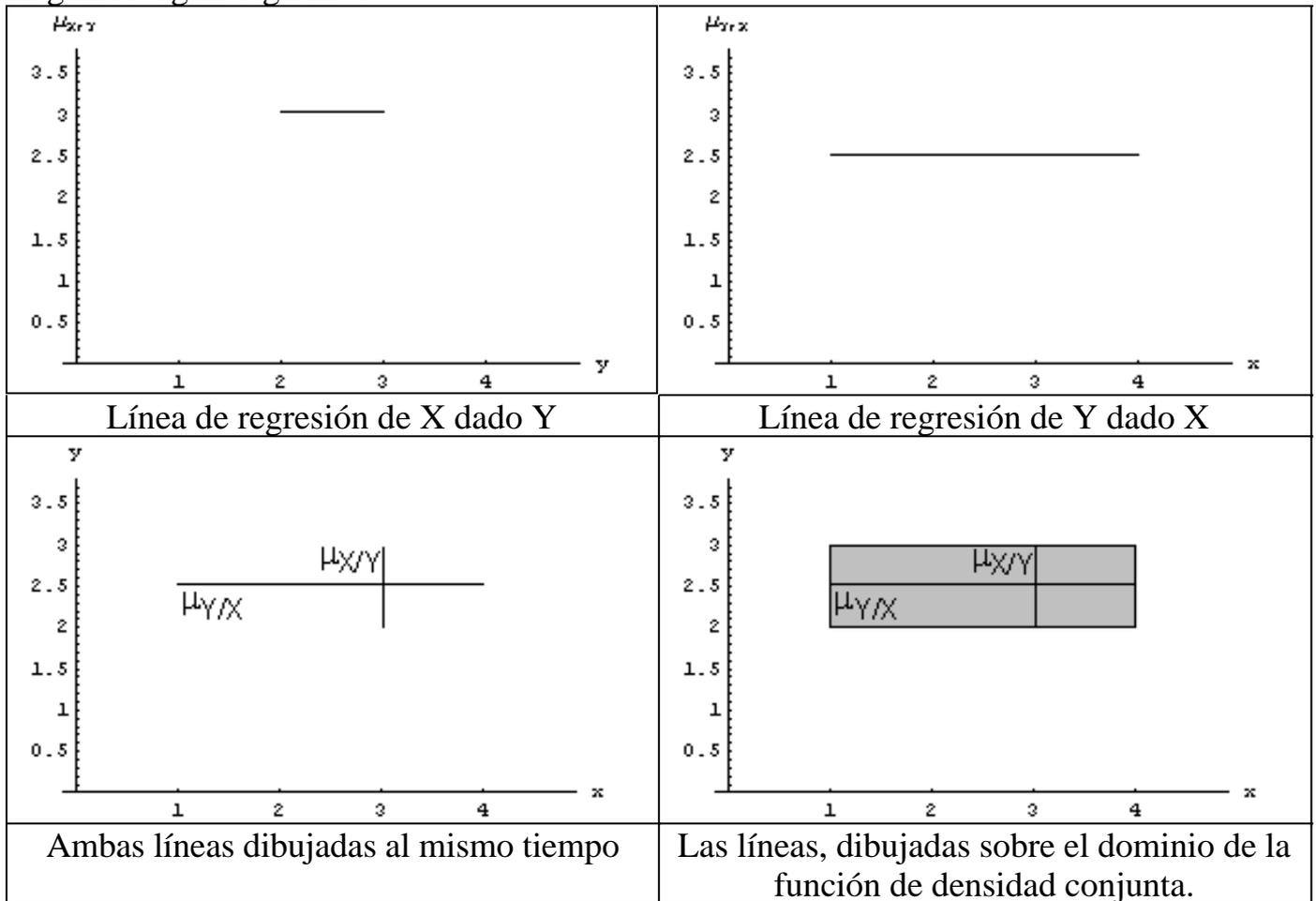
Luego, las esperanzas condicionales son:

$$\mu_{X/Y} = \int_{-\infty}^{\infty} x f_{X/Y}(x, y) dx = \int_{-\infty}^{\infty} x f_X(x) dx = \int_1^4 \frac{1}{21} x^2 dx = 3.04$$

$$\mu_{Y/X} = \int_{-\infty}^{\infty} y f_{Y/X}(x, y) dy = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_2^3 \frac{2}{5} y dy = 2.53$$

Vemos que en este caso, como las variables son independientes, las esperanzas

condicionales son constantes. De hecho, son directamente μ_x y μ_y respectivamente. Hagamos algunos gráficos:



Observemos que cuando las variables son independientes, las líneas de regresión son rectas y sin pendiente, y se cortan ortogonalmente.

Distribuciones discretas

La fórmula para calcular la esperanza condicional de X dado Y, para X e Y discretas, es análoga a la fórmula para las continuas:

$$E(X/Y) = \mu_{X/Y} = \sum_{x=-\infty}^{\infty} x P_{X/Y}(x, y)$$

Con respecto a las líneas de regresión, para variables discretas obviamente no existen. A cambio, se puede trazar un gráfico discreto de los valores de $\mu_{x|y}$ para los posibles valores de Y.

Problemas típicos

1) Se tienen las variables aleatorias continuas X e Y, cuya distribución

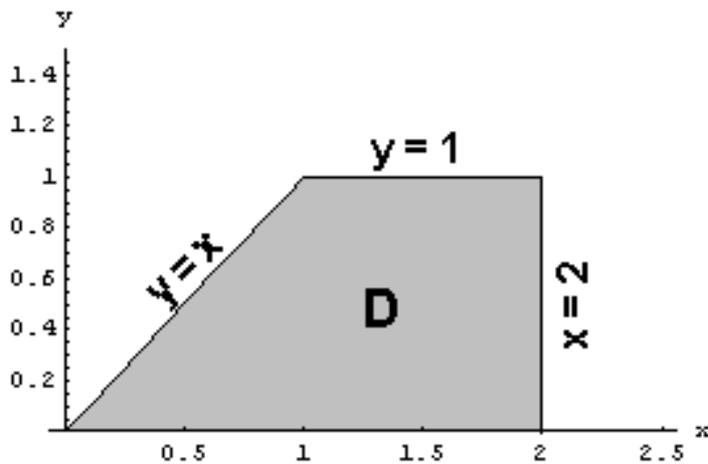
conjunta es la siguiente:

$$f_{XY}(x, y) = \begin{cases} \frac{8xy}{7} & (x, y) \in D \\ 0 & (x, y) \notin D \end{cases}$$

Se pide:

- Hallar las esperanzas condicionales
- Trazar las líneas de regresión

Resolución



a) Comenzamos por graficar la función de densidad para orientarnos en las cuentas que vamos a hacer.

Hallamos las distribuciones marginales:

$$f_X(x) = \begin{cases} \frac{4x^3}{7} & 0 < x < 1 \\ \frac{4x}{7} & 1 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{4}{7} y(4 - y^2) & 0 < y < 1 \\ 0 & \forall \text{ otro } y \end{cases}$$

Hallamos las distribuciones condicionales:

$$f_{X/Y}(x, y) = \begin{cases} \frac{2x}{4 - y^2} & y < x < 2, \quad 0 < y < 1 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

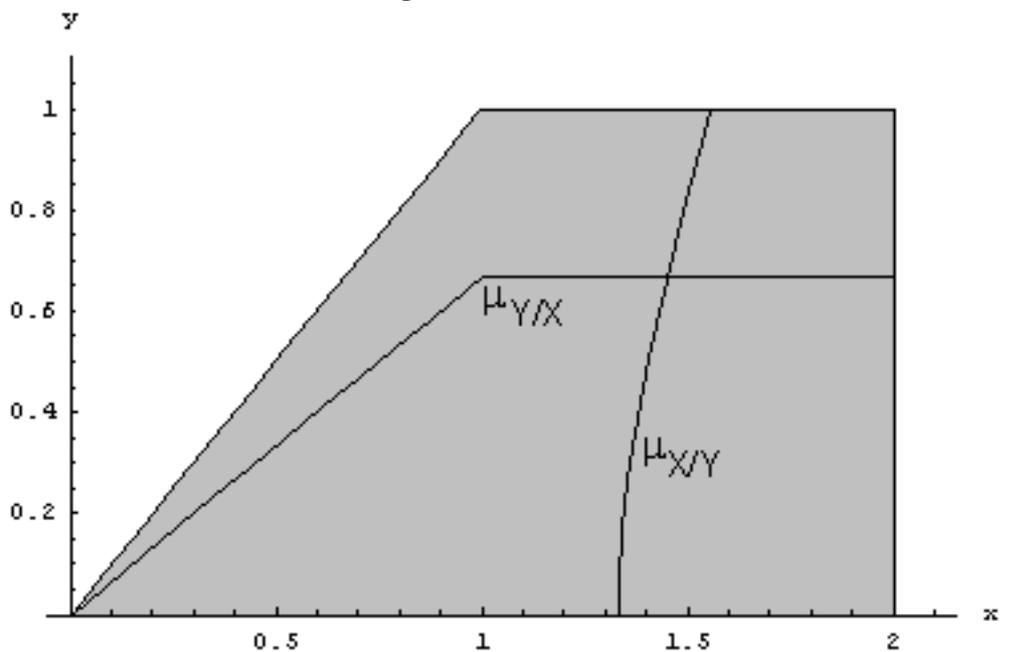
$$f_{Y/X}(x, y) = \begin{cases} \frac{2y}{x^2} & 0 < x < 1, \quad 0 < y < x \\ 2y & 1 < x < 2, \quad 0 < y < 1 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Hallamos las esperanzas condicionales

$$\mu_{X/Y} = \int_{-\infty}^{\infty} x f_{X/Y}(x, y) dx = \int_y^2 \frac{2x^2}{4-y^2} dx = \frac{2y^2 + 4y + 8}{3y + 6}$$

$$\mu_{Y/X} = \int_{-\infty}^{\infty} y f_{Y/X}(x, y) dy = \begin{cases} \int_0^x \frac{2y^2}{x^2} dy & 0 < x < 1 \\ \int_0^1 2y^2 dy & 1 < x < 2 \end{cases} = \begin{cases} \frac{2}{3}x & 0 < x < 1 \\ \frac{2}{3} & 1 < x < 2 \end{cases}$$

b) Trazamos las líneas de regresión



2) Se tienen las variables aleatorias discretas X e Y, cuya distribución conjunta es:

		Y		
		0	2	4
X	1	0.25	0.05	0.3
	2	0.15	0.1	0.15

Halle las esperanzas condicionales.

Resolución

Agregamos en los márgenes de la conjunta las distribuciones marginales:

		Y			P_x
X	P_{xy}	0	2	4	
	1	0.25	0.05	0.3	0.6
	2	0.15	0.1	0.15	0.4
P_y	0.4	0.15	0.45		

Hallamos las distribuciones condicionales:

		Y		
X	P_{xy}	0	2	4
	1	5/8	1/3	2/3
	2	3/8	2/3	1/3

		Y		
X	P_{y/x}	0	2	4
	1	5/12	1/12	6/12
	2	3/8	2/8	3/8

Hallamos las esperanzas condicionales:

$$\mu_{X/Y} = \sum_{x=-\infty}^{\infty} x P_{X/Y}(x, y) = \begin{cases} 1\frac{5}{8} + 2\frac{3}{8} & y=0 \\ 1\frac{1}{3} + 2\frac{2}{3} & y=2 \\ 1\frac{2}{3} + 2\frac{1}{3} & y=4 \end{cases} = \begin{cases} 1.375 & y=0 \\ 1.667 & y=2 \\ 1.333 & y=4 \end{cases}$$

$$\mu_{Y/X} = \sum_{y=-\infty}^{\infty} y P_{Y/X}(x, y) = \begin{cases} 0\frac{5}{12} + 2\frac{1}{12} + 4\frac{6}{12} & x=1 \\ 0\frac{3}{8} + 2\frac{2}{8} + 4\frac{3}{8} & x=2 \end{cases} = \begin{cases} 2.17 & x=1 \\ 2 & x=2 \end{cases}$$



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/process/text/>



El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandrod@probabilidad.com.ar>

Versión Actualizada al: 13 de mayo de 2004

Esperanza, Covarianza y Correlación

Ahora vamos a estudiar la esperanza y la varianza en más de una dimensión y a desarrollar nuevos conceptos.

Esperanza en más de una dimensión

Para una dimensión vimos que la esperanza de una función $\Phi(x)$ es:

$$E(Y) = \int_{-\infty}^{+\infty} \Phi(x) f_X(x) dx$$

Esto se puede generalizar. Para n variables, la esperanza de una variable Z definida a partir de una función como $Z = \Phi(x_1, x_2, \dots, x_n)$ es:

$$E(Z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \Phi(x_1, x_2, \dots, x_n) f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) dx_n \dots dx_2 dx_1$$

Luego, para dos variables, la esperanza de una función $Z = \Phi(x, y)$ es:

$$E(Z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(x, y) f_{XY}(x, y) dy dx$$

Estas y el resto de las fórmulas son análogas para el caso de variables discretas.

Esperanza de la suma de variables aleatorias

Sea $Z = X + Y$, es decir, $\Phi(x, y) = x + y$. Según acabamos de ver:

$$E(Z) = E(X + Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f_{XY}(x, y) dy dx$$

Como la integral de la suma es la suma de las integrales, hacemos:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f_{XY}(x, y) dy dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_{XY}(x, y) dy dx + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f_{XY}(x, y) dy dx$$

Hacemos la primera integral:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_{XY}(x, y) dy dx = \int_{-\infty}^{+\infty} x \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx = \int_{-\infty}^{+\infty} x f_X(x) dx = E(X)$$

Análogamente, la otra integral es:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f_{XY}(x, y) dy dx = E(Y)$$

Luego, llegamos a que:

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

De hecho, se puede demostrar análogamente que:

$$E\left(\sum_{i=1}^r a_i X_i\right) = \sum_{i=1}^r a_i E(X_i)$$

Es decir, que la esperanza de una combinación lineal de variables aleatorias es la combinación lineal de las esperanzas. Por eso se dice que la esperanza es un **operador lineal**.

Varianza de la suma de variables aleatorias

Sea $Z = X + Y$, es decir, $\Phi(x, y) = x + y$.

$$\begin{aligned} \sigma_Z^2 &= E((Z - E(Z))^2) = E((X + Y - E(X + Y))^2) = E((X + Y - (\mu_X + \mu_Y))^2) = \\ &= E((X + Y - \mu_X - \mu_Y)^2) = E(((X - \mu_X) + (Y - \mu_Y))^2) \end{aligned}$$

Desarrollando el cuadrado queda:

$$E((X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y))$$

Usando la linealidad del operador esperanza:

$$E((X - \mu_X)^2) + E((Y - \mu_Y)^2) + 2E((X - \mu_X)(Y - \mu_Y))$$

Reconocemos en la fórmula a las varianzas de X y de Y:

$$\sigma_X^2 + \sigma_Y^2 + 2E((X - \mu_X)(Y - \mu_Y))$$

A la esperanza que queda en el último término la llamaremos covarianza de X e Y, y la notaremos σ_{XY} .

En conclusión, llegamos a que:

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$$

Donde σ_{XY} es la covarianza de X e Y.

Vemos que **la varianza de la suma no es la suma de las varianzas**, porque aparece sumando la covarianza.

Antes de estudiar la covarianza, generalizamos la fórmula para la varianza de la combinación lineal de dos variables aleatorias:

$$\sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$$

(la demostración es similar a la que hicimos, pero agregando los coeficientes)

La covarianza

Según vimos, la covarianza de X e Y vale:

$$\sigma_{XY} = E((X-\mu_X)(Y-\mu_Y))$$

Observemos que $(X-\mu_X)(Y-\mu_Y)$ es una función de X e Y. Luego, podemos hallar su esperanza con la fórmula que dimos al principio, y queda:

$$\sigma_{XY} = E((X-\mu_X)(Y-\mu_Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x-\mu_X)(y-\mu_Y) f_{XY}(x, y) dy dx$$

Otra forma de calcularla es desarrollando el producto:

$$\sigma_{XY} = E((X-\mu_X)(Y-\mu_Y)) = E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y)$$

Usando la linealidad del operador esperanza:

$$E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y$$

Como $E(X) = \mu_X$ y $E(Y) = \mu_Y$:

$$E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y$$

Es decir, queda:

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y$$

Tenemos entonces dos formas posibles de calcular la covarianza, que podremos elegir según la ocasión:

$$\text{cov}(X, Y) = \sigma_{XY} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x-\mu_X)(y-\mu_Y) f_{XY}(x, y) dy dx = E(XY) - \mu_X\mu_Y$$

Interpretación de la covarianza

Notemos primeramente la similitud entre la fórmula de la varianza y la fórmula de la covarianza. Las fórmulas para calcular la varianza son:

$$\sigma_W^2 = \int_{-\infty}^{+\infty} (w-\mu_W) f_W(w) dw = E(W^2) - \mu_W^2$$

Vemos que si hacemos $X = Y = W$ en las fórmulas de la covarianza de X e Y, llegamos a fórmulas similares a las de la varianza para una sola variable. Esto nos

dice que de alguna forma la covarianza mide la relación **entre** dos variables aleatorias, es decir, no la variabilidad de una variable respecto de sí misma, como la varianza, sino la relación entre dos variables distintas.

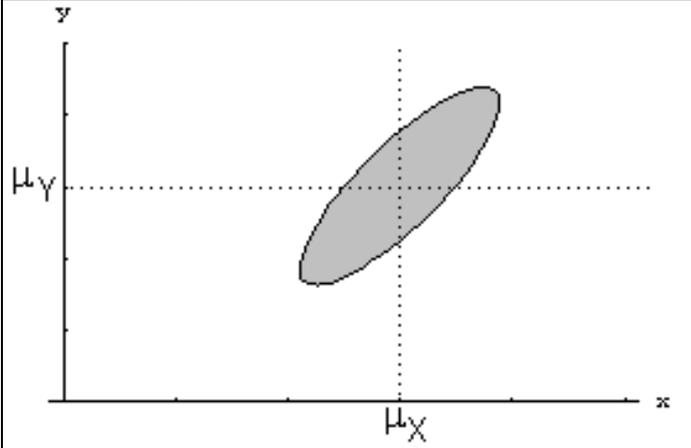
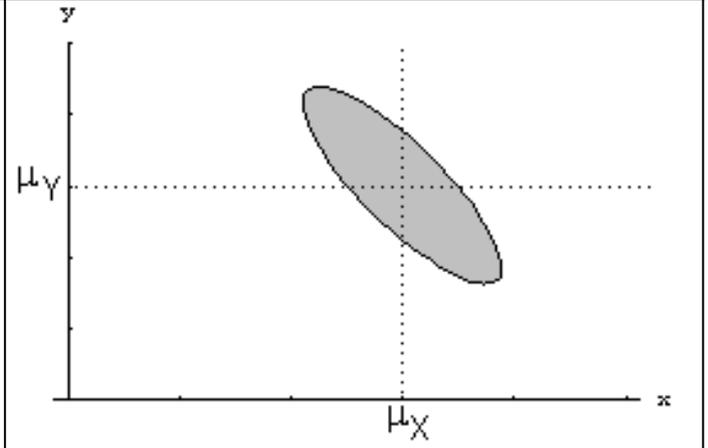
Observemos la fórmula:

$$\sigma_{XY} = E((X - \mu_X)(Y - \mu_Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y) f_{XY}(x, y) dy dx$$

Cuando X es grande e Y es grande, el producto $(X - \mu_X)(Y - \mu_Y)$ será positivo, y también será positivo cuando X es chica e Y es chica. Cuando una es grande y la otra es chica, el producto $(X - \mu_X)(Y - \mu_Y)$ será negativo.

Como la integral es la suma de esos infinitos productos diferenciales, dará positiva si la distribución conjunta asigna más densidad a las zonas donde X e Y son las dos chicas o las dos grandes, y menos densidad a las zonas donde una es grande y la otra es chica.

Veamos un gráfico:

	
<p>Esta distribución se compone predominantemente de valores en los cuales X e Y son ambas grandes o ambas chicas. La covarianza dará un valor positivo.</p>	<p>Esta distribución se compone predominantemente de valores en los cuales X e Y son una grande y la otra chica. La covarianza dará un valor negativo.</p>

El coeficiente de correlación lineal

Mirando los gráficos, observamos que el de la izquierda se asemeja a una función lineal positiva, y el de la izquierda se asemeja a una función lineal negativa. Es decir, las distribuciones ilustradas dan la impresión de estar ubicadas sobre una recta.

Si la distribución que tenemos fuese como en alguno de los dos casos ilustrados, podríamos suponer que existe algún tipo de relación lineal entre las variables.

Podemos medir qué tan "lineal" es una distribución, mediante el coeficiente de correlación. Dicho coeficiente se nota con la letra ρ , y vale:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

El coeficiente de correlación lineal asume valores entre -1 y 1. Si es positivo, quiere decir que hay una relación lineal positiva. Si es negativo, quiere decir que hay una relación lineal negativa. Cuanto más cercano a 1 o -1 sea, quiere decir que más fuerte es la relación lineal. Si está cercano a cero, quiere decir que no hay una relación lineal.

Veamos algunos ejemplos:

<p>ρ será un número positivo</p>	<p>ρ será un número muy cercano a 1</p>	<p>ρ será un número cercano a 0</p>

Variables aleatorias independientes

¿Cómo se modifica todo lo que dijimos hasta ahora cuando las variables son independientes?

Cuando las variables son independientes, no hay ningún tipo de relación entre ellas, ni lineal, ni ninguna otra cosa. **Luego, la covarianza de dos variables independientes es cero.**

Se demuestra fácilmente. Veamos:

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y$$

donde:

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{XY}(x, y) dy dx$$

Si las variables son independientes, $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$. Luego:

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_X(x) f_Y(y) dy dx = \int_{-\infty}^{+\infty} x f_X(x) \int_{-\infty}^{+\infty} y f_Y(y) dy dx = \int_{-\infty}^{+\infty} x f_X(x) \mu_Y dx = \\ &= \mu_Y \int_{-\infty}^{+\infty} x f_X(x) dx = \mu_X \mu_Y \end{aligned}$$

Luego como $E(XY) = \mu_X \mu_Y$, obtenemos que $\sigma_{XY} = 0$.

Demostramos que si dos variables son independientes, su covarianza es cero. Hagamos sin embargo una observación: la recíproca no es válida. Es decir, el hecho de que la covarianza sea cero no implica que las variables sean

independientes. Solamente implica que no tienen relación lineal (pero puede haber otros tipos de relaciones).

Consecuencias de que la covarianza sea cero:

- El coeficiente de correlación es cero.
- La varianza de la combinación lineal de dos variables es:

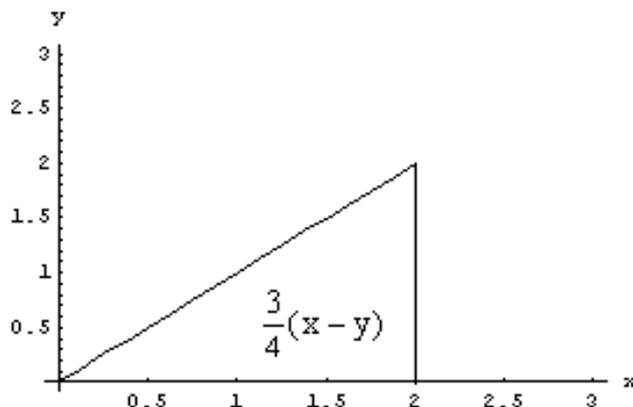
$$\sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$$

- Generalizando, la varianza de la combinación lineal de n variables independientes es:

$$\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i)$$

Problemas típicos

1) Sean las variables aleatorias continuas X e Y cuya función de densidad conjunta es:



$$f_{XY}(x, y) = \begin{cases} \frac{3}{4}(x-y) & 0 < x < 2, 0 < y < x \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Halle la esperanza de $Z = XY^2$

Resolución

Como X e Y son continuas, la esperanza de una función $Z = \Phi(x, y)$ es:

$$E(Z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(x, y) f_{XY}(x, y) dy dx$$

Luego:

$$E(XY^2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy^2 f_{XY}(x, y) dy dx = \int_0^2 \int_0^x xy^2 \frac{3}{4}(x-y) dy dx = \frac{2}{3}$$

2) Sean las variables aleatorias discretas X e Y, cuya distribución conjunta es:

		Y		
X	P_{XY}	1	2	3
	1	0.12	0.1	0.08
	2	0.28	0.2	0.22

Halle la esperanza de $Z = XY^2$

Resolución

Como X e Y son discretas, la esperanza de una función $Z = \varphi(x,y)$ es:

$$E(Z) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} \varphi(x, y) P_{XY}(x, y)$$

Luego:

$$E(XY^2) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} xy^2 P_{XY}(x, y) =$$

$$= 1 \cdot 1^2 \cdot 0.12 + 1 \cdot 2^2 \cdot 0.1 + 1 \cdot 3^2 \cdot 0.08 + 2 \cdot 1^2 \cdot 0.28 + 2 \cdot 2^2 \cdot 0.2 + 2 \cdot 3^2 \cdot 0.22 = 7.36$$

3) Halle la esperanza y la varianza de la suma de las variables del problema 2.

Resolución

a) Para calcular la esperanza de la suma, podemos proceder de dos maneras:

• Por definición, haciendo:

$$E(X + Y) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} (x + y) P_{XY}(x, y) =$$

$$= 2 \cdot 0.12 + 3 \cdot 0.1 + 4 \cdot 0.08 + 3 \cdot 0.28 + 4 \cdot 0.2 + 5 \cdot 0.22 = 3.6$$

• Con la propiedad de que la esperanza de la suma es la suma de las esperanzas:

Hallamos primeramente las distribuciones marginales:

		Y			P_X
X	P_{XY}	1	2	3	
	1	0.12	0.1	0.08	0.3
	2	0.28	0.2	0.22	0.7
P_Y		0.4	0.3	0.3	---

Calculamos las esperanzas de X e Y:

$$E(X) = 1 \cdot 0.3 + 2 \cdot 0.7 = 1.7$$

$$E(Y) = 1 \cdot 0.4 + 2 \cdot 0.3 + 3 \cdot 0.3 = 1.9$$

$$\text{Luego } E(X+Y) = E(X) + E(Y) = 1.7 + 1.9 = 3.6$$

b) Para calcular la varianza de la suma, vamos a usar la fórmula:

$$\sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$$

$$\sigma_X^2 = E(X^2) - E(X)^2$$

$$E(X^2) = 1 \cdot 0.3 + 2^2 \cdot 0.7 = 3.1$$

$$\sigma_X^2 = 3.1 - 1.7^2 = \mathbf{0.21}$$

$$\sigma_Y^2 = E(Y^2) - E(Y)^2$$

$$E(Y^2) = 1^2 \cdot 0.4 + 2^2 \cdot 0.3 + 3^2 \cdot 0.3 = 4.3$$

$$\sigma_Y^2 = 4.3 - 1.9^2 = \mathbf{0.69}$$

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y$$

$$E(X+Y) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} (x+y) P_{XY}(x, y) =$$

$$= 1 \cdot 0.12 + 2 \cdot 0.1 + 3 \cdot 0.08 + 2 \cdot 0.28 + 4 \cdot 0.2 + 6 \cdot 0.22 = 3.24$$

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = 3.24 - 3.23 = 0.01$$

Y en este caso, $a = b = 1$. Luego:

$$\sigma_{X+Y}^2 = 1^2\sigma_X^2 + 1^2\sigma_Y^2 + 2\sigma_{XY} = 0.21 + 0.69 + 0.02 = 0.92$$

4) Halle la esperanza, la varianza y el coeficiente de correlación lineal de la suma de las variables del problema 1.

Resolución

A partir de la conjunta calculamos:

$$f_X(x) = \begin{cases} \frac{3}{8}x^2 & 0 < x < 2 \\ 0 & \forall \text{ otro } x \end{cases} \quad f_Y(y) = \begin{cases} \frac{3}{2} \left(\frac{1}{4}y^2 - y + 1 \right) & 0 < y < 2 \\ 0 & \forall \text{ otro } y \end{cases}$$

$$\mu_X = 3/2 \quad ; \quad \mu_Y = 1/2$$

$$\sigma_X^2 = 3/20 \quad ; \quad \sigma_Y^2 = 3/20$$

$$\sigma_{XY} = 4/5 - 3/4 = 1/20$$

Luego:

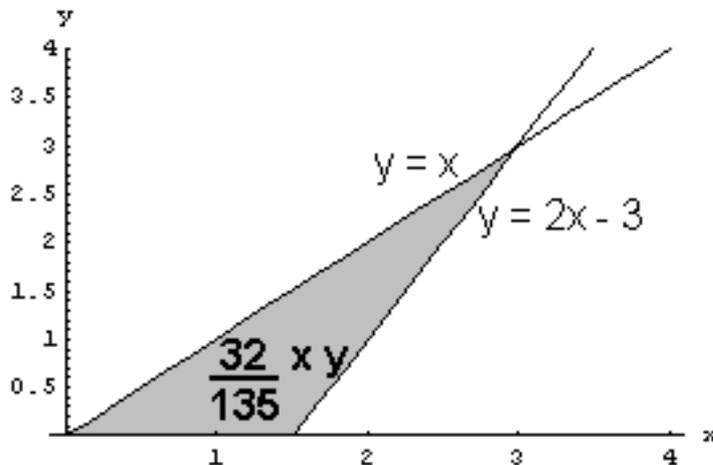
$$E(X+Y) = E(X) + E(Y) = 3/2 + 1/2 = 2$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} = \frac{3}{20} + \frac{3}{20} + \frac{2}{20} = \frac{8}{20}$$

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{1}{20} \frac{1}{\sqrt{\frac{3}{20}} + \sqrt{\frac{3}{20}}} = 0.0645$$

El coeficiente de correlación lineal arrojó prácticamente el valor cero. Esto nos dice que no hay relación lineal

5) Halle la esperanza, la varianza y el coeficiente de correlación lineal de la suma de las variables X e Y cuya función de densidad conjunta es:



$$f_{XY}(x, y) = \begin{cases} \frac{32}{135}xy & y < x < \frac{y+3}{2}, 0 < y < 3 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Resolución

Calculamos:

$$\mu_X = 51/25 \quad ; \quad \mu_Y = 42/25$$

$$\sigma_X^2 = 621/2500 \quad ; \quad \sigma_Y^2 = 261/625$$

$$\sigma_{XY} = 93/25 - 2142/625 = 183/625$$

Luego:

$$E(X+Y) = E(X) + E(Y) = 51/25 + 42/25 = 93/25$$

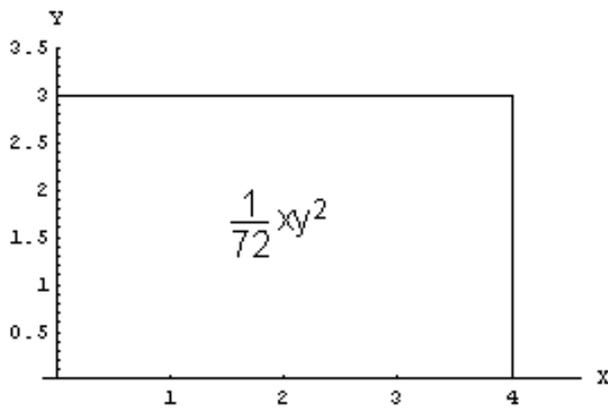
$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} = \frac{621}{2500} + \frac{261}{625} + 2\frac{183}{625} = \frac{3129}{2500}$$

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{183}{\sqrt{\frac{621}{2500} + \frac{261}{625}}} = 0.26$$

El coeficiente de correlación lineal arrojó un valor positivo, lo suficientemente lejano del cero como para ser tenido en cuenta. Esto nos dice que existe una relación lineal positiva entre X e Y. Además, como el valor tampoco es cercano al uno, dicha relación lineal es débil.

Esto resulta absolutamente compatible con lo que nos muestra el gráfico, donde el dominio es "parecido" a un segmento de recta de pendiente positiva.

6) Halle la esperanza, la varianza y el coeficiente de correlación lineal de la suma de las variables X e Y cuya función de densidad conjunta es:



$$f_{XY}(x, y) = \begin{cases} \frac{1}{72}xy^2 & 0 < x < 4, 0 < y < 3 \\ 0 & \forall \text{ otro } x, y \end{cases}$$

Resolución

Calculamos:

$$\mu_X = 8/3 \quad ; \quad \mu_Y = 9/4$$

$$\sigma_X^2 = 8/9 \quad ; \quad \sigma_Y^2 = 27/80$$

$$\sigma_{XY} = 6 - 6 = 0$$

(aunque si nos hubiéramos dado cuenta de que las variables son independientes, habríamos sabido que la covarianza es cero sin calcularla).

Luego:

$$E(X+Y) = E(X) + E(Y) = 51/25 + 42/25 = 59/12$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} = \frac{8}{9} + \frac{27}{80} + 0 = \frac{883}{720}$$

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{0}{\sqrt{\frac{8}{9}} + \sqrt{\frac{27}{80}}} = 0$$

Nuevamente, no hacía falta hacer la cuenta para saber que el coeficiente de correlación lineal, puesto que si las variables son independientes, se deduce que no hay relación lineal (ni de ningún otro tipo).

7) Halle la esperanza y la varianza de $Z = 3X+2Y$ con las variables X e Y tomadas del problema 1.

Resolución

Este problema es como el 4 pero más general, porque no es una suma sino una combinación lineal, es decir X e Y vienen acompañados por coeficientes distintos de uno.

Vamos a usar las fórmulas generales, y los datos que ya calculamos en el problema 4.

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i) = 3E(X) + 2E(Y) = 3 \cdot \frac{3}{2} + 2 \cdot \frac{1}{2} = \frac{11}{2}$$

$$\sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY} = 3^2 \cdot \frac{3}{20} + 2^2 \cdot \frac{3}{20} + 2 \cdot 3 \cdot 2 \cdot \frac{1}{20} = \frac{51}{20}$$

8) La distribución del peso en kg de una sandía es:

$$f_X(x) = \begin{cases} 1 & 1,5 < x < 2,5 \\ 0 & \forall \text{ otro } x \end{cases}$$

Se pide:

- Las sandías se venden a \$3/kg. ¿Cuál es la media y la varianza del precio de venta de una sandía?
- Si se colocan 3 sandías en una bolsa, ¿Cuál es la media y la varianza del peso de la bolsa?
- Compare los resultados de a) y b) y extraiga conclusiones.

Resolución

Vamos a necesitar la media y la varianza de X:

$$E(X) = 2$$

$$\sigma_X^2 = 1/12$$

a) $Y = 3X$. Luego, por las propiedades estudiadas en el capítulo 2:

$$E(Y) = E(aX) = a E(X) = 3 E(X) = 3 \cdot 2 = 6$$

$$\sigma_Y^2 = \sigma_{aX}^2 = a^2 \sigma_X^2 = 9 \cdot 1/12 = 3/4$$

b) Aquí podemos equivocarnos y volver a plantear $Y = 3X$. Pero eso está mal, porque sería tomar el peso de una sandía y multiplicarlo por 3, cuando en realidad son 3 sandías distintas. En el punto c analizaremos la incidencia de este error. Lo correcto es que $Y = X_1 + X_2 + X_3$, donde X_1, X_2 y X_3 son 3 variables distintas e independientes, y las 3 tienen la distribución que figura en el enunciado.

Por las propiedades que hemos estudiado:

$$E(Y) = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 2 + 2 + 2 = 6$$

Vemos que la esperanza dio igual que en a).

Para la varianza de una combinación lineal de variables independientes, tenemos:

$$Var\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 Var(X_i) = 1^2 \sigma_{X_1}^2 + 1^2 \sigma_{X_2}^2 + 1^2 \sigma_{X_3}^2 = \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = \frac{3}{12} = \frac{1}{4}$$

Vemos que la varianza nos dio menor que en a).

c) Vemos que las medias nos quedaron iguales, pero la varianza de la suma de 3 sandías quedó menor que la varianza de multiplicar una sandía por 3. ¿Por qué sucede esto? La varianza es una medida de cuánto tienden a alejarse de la media los valores de la variable aleatoria. Como los valores pueden estar a la izquierda o a la derecha de la media, sumando n variables aleatorias las distancias a la media de

cada uno de esos n valores tienden a compensarse, por eso la varianza de la suma de n variables es menor que la de la multiplicación de una variable por n . Por este motivo, para llegar a resultados correctos, es importante en todos los problemas darse cuenta de cuándo se está multiplicando una variable por un coeficiente, y cuándo se están sumando variables distintas (que pueden ser, por ejemplo, independientes e idénticamente distribuidas).



ABC Amber Text Converter Trial version

Please register to remove this banner.

<http://www.thebeatlesforever.com/processtext/>

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 12 de julio de 2004

CAPÍTULO IV

Proceso de Bernoulli

Experimento de Bernoulli

Es un experimento que puede arrojar 2 resultados posibles. A uno de los resultados se lo denomina arbitrariamente "éxito" y al otro "fracaso". El experimento de Bernoulli lleva asociada una probabilidad (la probabilidad de "éxito").

Veamos el ejemplo siguiente:

Ejemplo

Si voy a tirar un dado, y lo que voy a observar es si sale o no sale un 5, entonces esto puede ser visto como un experimento de Bernoulli constituido así:

- Éxito: que salga un 5
- Fracaso: que no salga un 5
- Probabilidad de éxito: $p = 1/6$
- Probabilidad de fracaso: $q = 1-p = 5/6$

En ese ejemplo vemos que llamamos "éxito" a que salga un 5, porque justamente estábamos observando si iba a salir o no un 5. El hecho de llamar a algo "éxito" o "fracaso" no tiene nada que ver con que sea "bueno" o "malo" respectivamente, sino con el hecho de que haya dado positiva o negativa la observación que queríamos hacer.

Como vimos, p es la probabilidad de éxito, es decir, la probabilidad de que se cumpla la condición que queríamos observar. Y la probabilidad de fracaso, es decir, de no-éxito, $1-p$, a menudo se encuentra escrita como q .

Proceso de Bernoulli

Consiste en hacer n veces un experimento de Bernoulli, teniendo en cuenta:

- que las condiciones no varían. (Ejemplo: la moneda que arrojé n veces sigue siendo la misma y no se deforma). Es decir, que la probabilidad p de obtener un éxito en la 5ta vez es la misma que la de obtener un éxito en la 8va vez.

- que cada uno de los experimentos es independiente (Ejemplo: que haya salido cara en la 5ta vez que tiré la moneda, no me afecta lo que salga en la 8va vez).

Se definen las siguientes variables:

- n : la cantidad de veces que se hace el experimento
- p : la probabilidad de que un experimento arroje éxito.
- k : la cantidad de veces que se obtiene éxito en las n veces que se hace el experimento.

Ejemplo

Si arrojó una moneda 8 veces, con probabilidad 0,5 de que salga cara (considerando cara como éxito) y sale cara 5 veces, tengo:

- $n = 8$
- $p = 0,5$
- $k = 5$

Generalmente conocemos el valor de p , y entonces nos preguntamos cuántos éxitos obtendremos haciendo el experimento una determinada cantidad de veces, o cuántas veces tendremos que hacer el experimento para obtener una determinada cantidad de éxitos.

De esta forma obtenemos 2 distribuciones:

- Binomial: consiste en preguntar por la cantidad de éxitos en n veces. Es decir, dado n , calcular la distribución de k .
- Pascal: consiste en preguntar por la cantidad de veces necesarias para obtener k éxitos. Es decir, dado k , calcular la distribución de n .

Y además:

- Geométrica: caso particular de Pascal cuando $k = 1$, es decir, consiste en preguntar por la cantidad de veces necesarias para obtener el primer éxito.

Distribución Binomial

"¿Cuál es la probabilidad de obtener x éxitos en n intentos?"

Si

X:Bi (n ; p)

es decir: X es una variable binomial con parámetros n y p

es decir: X es la variable que representa la cantidad de éxitos obtenidos en n experimentos de Bernoulli independientes cada uno con probabilidad de éxito

p

$$P(X = x) = \begin{cases} \binom{n}{x} p^x \cdot (1-p)^{n-x} & 0 \leq x \leq n \\ 0 & \forall \text{ otro } x \end{cases}$$

entonces:

$$E(X) = n \cdot p$$

$$\sigma^2_X = n \cdot p \cdot (1-p)$$

n es un número natural

p es un número real entre 0 y 1

Propiedades reproductivas

Si tenemos

- m variables X_i
- $X_i: \text{Bi}(n_i, p)$
- X_i independiente de X_j para $i \neq j$

$$Y = \sum_{i=1}^m X_i$$

•

entonces:

- $Y: \text{Bi}(n_Y, p)$

$$n_Y = \sum_{i=1}^m n_i$$

•

Es decir, la suma de m variables binomiales independientes cada una con igual p y con su propio n resulta ser una variable binomial con el mismo p que las anteriores y n dado por la suma de los n de las variables originales.

Estrategia

Sabemos que nos encontramos frente a la necesidad de emplear una distribución binomial cuando:

- nos dan una determinada cantidad de elementos (piezas, intentos, etc.)
- cada uno de esos elementos puede o no cumplir con una determinada condición (que la pieza sea defectuosa, que el intento haya salido bien, etc.)
- nos dan o es posible calcular la probabilidad de que **un** elemento cumpla con la condición
- nos preguntan cuál es la probabilidad de que determinada cantidad de elementos, de los n que hay en total, cumplan con la condición.

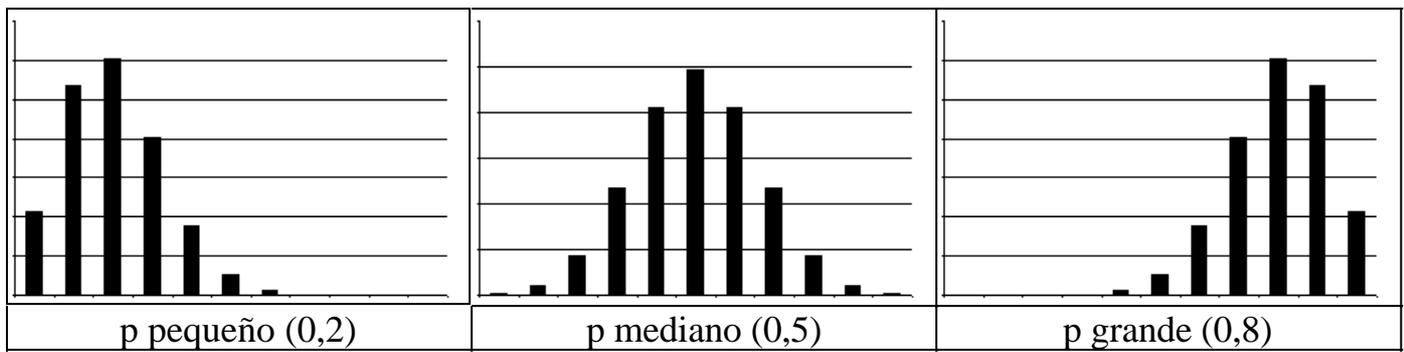
Por lo general estos problemas se resuelven encontrando la forma de calcular la probabilidad de que **un** elemento cumpla con la condición sin importar cuántos elementos haya. Luego tomaremos una variable X que representará cuántos elementos de los n que hay en total cumplen con la condición. Sus parámetros serán:

- p : la probabilidad de que **un** elemento cumpla con la condición
- n : la cantidad de elementos que hay en total.

Siempre comenzaremos por suponer que los n elementos son independientes entre sí, es decir, que el hecho de que un elemento cumpla o no con la condición no afecta la probabilidad de que los demás la cumplan o no. De lo contrario no podríamos usar la distribución binomial porque no estaríamos cumpliendo con las características del proceso de Bernoulli.

Si X está distribuida binomialmente con n y p , $P(X = x)$ tendrá valor no nulo $\forall x \in [0 ; n]$. Todos los demás x tienen probabilidad nula. De todas las distribuciones que estudiaremos, ésta es la única que está acotada tanto superior como inferiormente.

Aspecto



Vemos que todos los valores entre 0 y n tienen probabilidad no nula, aunque la probabilidad de los valores cercanos a n será muy pequeña si p es chico, y la probabilidad de los valores cercanos al 0 será muy pequeña si p es grande.

Problemas típicos

1) ¿Cuál es la probabilidad de obtener cara 5 veces al arrojar una moneda 8 veces?

Resolución:

Comenzaremos por asumir:

- que la moneda no es cargada (es decir, que hay probabilidad 0,5 de que salga cara, y 0,5 de que salga ceca)
- que la moneda conserva sus propiedades durante todo el proceso (es decir, que $P(\text{cara})$ se mantiene constante).
- que los intentos son independientes (es decir, que salga cara en el 3er intento no afecta la probabilidad de salga o no salga cara en el 8vo intento).

Bajo esas hipótesis, si llamamos éxito al hecho de obtener cara al tirar la moneda, la cantidad de éxitos que obtendremos en 8 veces será una variable binomial con $n = 8$ y $p = 0,5$.

Si a esa cantidad la llamamos X, podemos escribir:

$X: Bi(n = 8 ; p = 0,5)$

Nos piden la probabilidad de obtener 5 caras, es decir, la probabilidad de que $X = 5$.

$$P(X = 5) = \text{comb}(n,x) \cdot p^x \cdot (1-p)^{n-x} = \text{comb}(8,5) \cdot 0,5^5 \cdot 0,5^3 = 0,21875$$

2) Una máquina produce un determinado tipo de piezas. Las piezas a veces salen defectuosas. La probabilidad de que una pieza salga defectuosa es 0,01.

- ¿Cuál es la probabilidad de que haya piezas defectuosas en un lote de 50 piezas?
- ¿Cuál es la probabilidad de que haya 2 ó más piezas defectuosas en dicho lote?

c) ¿Cuál es la cantidad esperada de piezas defectuosas en el lote?
¿Cuál es la varianza?

Resolución:

a) El lote está formado por 50 piezas. Supondremos que las 50 piezas son independientes, en el sentido de que el hecho de que una pieza sea o no defectuosa no afecta la probabilidad de que las otras lo sean o no. Si no asumiéramos esto, no cumpliríamos con las condiciones del proceso de Bernoulli, por lo cual no podríamos aplicar la distribución binomial.

X: cantidad de piezas defectuosas en el lote

=> X:Bi(n ; p)

con

$$p = 0,01$$

$$n = 50$$

Luego:

$$P(\text{haya piezas defectuosas en el lote}) = P(X > 0) = 1 - P(X \leq 0) = 1 - P(X = 0)$$

$$= 1 - \text{comb}(n;x) \cdot p^x \cdot (1-p)^{n-x} = 1 - \text{comb}(50;0) \cdot 0,01^0 \cdot 0,99^{50} = 1 - 0,60501 = 0,39499$$

b) Como las condiciones son las mismas, podemos seguir usando la misma variable aleatoria X que antes, y entonces:

$$P(\text{haya dos o más piezas defectuosas en el lote}) = P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0,60501 - 0,30556 = 0,08944$$

c) La media de una variable binomial es n.p y la varianza es n.p.q es decir n.p.(1-p)

$$E_x = n \cdot p = 0,5$$

$$\sigma^2_x = n \cdot p \cdot (1 - p) = 0,495$$

3) En una determinada ciudad, el 20% de las personas tiene el cabello rubio y el 80% tiene el cabello negro. En esa población, 6 de cada 10 personas son hombres. Tomando una persona al azar, existe una probabilidad 0,7 de que esa persona tenga ojos oscuros.

Si en un colectivo hay 20 personas, ¿cuál es la probabilidad de encontrar más de 2 mujeres rubias de ojos claros? ¿Qué suposiciones debe hacer para poder resolver el problema?

Resolución:

Este ejemplo lo que pretende es confundirnos con el cálculo del p, o bien desviar nuestra atención hacia la composición de la población para que no nos demos cuenta de que en realidad la pregunta es de naturaleza binomial.

Antes de comenzar, asumiremos que las personas son independientes y que el hecho de que la muestra sea tomada sobre un colectivo no afecta la composición.

Además tendremos que considerar infinita la cantidad de personas en la ciudad, pues de lo contrario las características de las personas según está planteado el problema ya no serían independientes. Es decir, si en la ciudad hubiera pocas

personas, el encontrar una persona de ojos claros en el colectivo haría más pequeña la probabilidad de encontrar otras personas de ojos claros en el colectivo.

Entonces comenzamos por hallar la probabilidad de que **una** persona cumpla con la condición, y luego usaremos la distribución binomial para trabajar con n personas.

Condiciones: mujer, rubia, ojos claros

Aquí tenemos que suponer que el sexo y el color de los cabellos y los ojos de una determinada persona también son independientes. Bajo esa suposición, podemos escribir:

$$P(\text{mujer} \cap \text{rubia} \cap \text{ojos claros}) = P(\text{mujer}) \cdot P(\text{rubia}) \cdot P(\text{ojos claros}) = 0,4 \cdot 0,2 \cdot 0,3 = 0,024.$$

Luego si tomamos X: cantidad de mujeres rubias de ojos claros en el colectivo

Tendremos que $X: Bi(n = 20 ; p = 0,024)$

$$\text{Luego } P(X > 2) = 1 - P(X \leq 2) = 1 - P(X = 0) - P(X = 1) - P(X = 2) = 0,01161$$

4) Se arrojan 3 dados sobre una mesa, y 4 dados sobre otra mesa. ¿Cuál la probabilidad de que no salga ningún 6?

Resolución:

Este ejemplo ilustra las propiedades reproductivas de la distribución binomial.

Podríamos tomar 2 variables binomiales, una para cada mesa, y entonces las dos tendrían probabilidad 5/6 y la primera tendría n = 3 y la segunda n = 4. Y luego las sumaríamos para obtener otra variable con la misma p, y n = 3 + 4 = 7. Pero esto es lo mismo que directamente considerar una sola variable para los 7 dados desde el principio, y de esto nos damos cuenta porque intuitivamente sabemos las propiedades reproductivas de la distribución binomial. Pero lo haremos de la primera forma, pues la manera de resolver la segunda ya ha sido mostrada en los ejemplos anteriores.

Tomamos:

$$\left. \begin{array}{l} X : Bi(n_x = 3; p = 5 / 6) \\ Y : Bi(n_y = 4; p = 5 / 6) \\ X, Y \text{ independientes} \\ Z = X + Y \end{array} \right\} Z : Bi(n_z = 7; p = 5 / 6)$$

$$P(\text{que no salga ningún 6}) = P(Z = 0) = 4 \cdot 10^{-6}$$

5) Si se tira una moneda una determinada cantidad de veces, se sabe que la cantidad de veces que sale cara es una variable binomial cuya media es 5 y su varianza es 2,5. ¿Diría Ud. que la moneda es honesta?

Resolución

Para que la moneda sea honesta, la probabilidad de que salga cara tiene que ser 0,5. Nos dicen que la moneda se tiró n veces, y que la cantidad de veces que salió cara

fue una variable binomial cuya media es 5 y su varianza es 2,5.

Entonces si X es esa variable binomial,

$$E_x = 5 \text{ y } \sigma_x^2 = 2,5$$

Con lo cual

$$\begin{cases} n p = 0.5 \\ n p (1 - p) = 2,5 \end{cases}$$

Nos queda un sistema de 2 ecuaciones con 2 incógnitas. Si lo resolvemos obtenemos:

$$n = 10$$

$$p = 0,5$$

Y como $p = 0,5$ concluimos que la moneda es honesta. (Y que se tiró 10 veces)

Distribución Geométrica

"¿Cuál es la probabilidad de obtener el primer éxito en el intento número x ?"

Si

X:Geom (p)

es decir: X es una variable geométrica con parámetro p

es decir: X es la variable que representa el número del intento en el cual se obtiene el primer éxito en experimentos de Bernoulli independientes cada uno con probabilidad de éxito p.

entonces:

$$P(X = x) = \begin{cases} p \cdot (1-p)^{x-1} & x \geq 1 \\ 0 & \forall \text{ otro } x \end{cases}$$

$$E(X) = \frac{1}{p}$$

$$\sigma_X^2 = \frac{1-p}{p^2}$$

p es un número real entre 0 y 1

Propiedades reproductivas

Si tenemos

- m variables X_i
- $X_i: \text{Geom}(p)$
- X_i independiente de X_j para $i \neq j$

$$Y = \sum_{i=1}^m X_i$$

•

entonces:

- $Y: \text{Pas}(k, p)$
- $k = m$

Es decir, la suma de m variables geométricas independientes, todas con igual p, resulta ser una variable de pascal con el mismo p que las anteriores y k dado por la

cantidad de variables geométricas que estamos sumando, es decir, m .

Estrategia

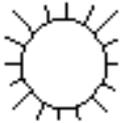
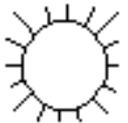
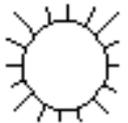
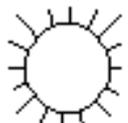
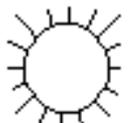
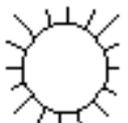
Sabemos que nos encontramos frente a una distribución geométrica cuando:

- nos dicen que vamos a repetir un determinado experimento hasta que logremos un éxito (ejemplo: que vamos a revisar piezas hasta que encontremos una que no sea defectuosa, o que vamos a disparar contra un blanco tantas veces como sea necesario hasta que acertemos, o que vamos a observar días hasta que haya un día soleado, etc.)
- nos dan o podemos calcular la probabilidad de tener éxito en cada uno de los intentos (la probabilidad de que cada pieza sea buena, la probabilidad de acertar cada vez que disparamos, la probabilidad de que un día sea soleado, etc.)
- nos preguntan cuál es la probabilidad de que logremos el objetivo en menos de x repeticiones, o la probabilidad de que nos tome más de x intentos lograr el objetivo, o la probabilidad de que lo logremos exactamente en el x -ésimo intento.

La única dificultad que esta distribución puede presentar es el cálculo de la probabilidad de tener éxito en cada uno de los intentos. Una vez obtenido ese valor, tendremos el parámetro p de la distribución, y el uso de la fórmula será inmediato.

La distribución geométrica en realidad es un caso particular de la distribución de Pascal (explicada en la siguiente sección). Una variable geométrica puede ser vista como una variable de Pascal cuyo parámetro p es el mismo que el de la geométrica, y cuyo parámetro k es igual a 1. De ahí que sumar variables geométricas es en esencia como sumar variables de Pascal, y de ahí que la suma de variables geométricas es una variable de Pascal. Por esto, si sospechamos que en un problema tendremos que sumar variables geométricas, puede resultar una idea bastante práctica considerarlas desde el principio variables de Pascal. De hecho la distribución geométrica se enseña separada de la pascal porque es más fácil aprender del caso particular al caso general.

Una característica de la distribución geométrica que es importante destacar, es lo que se conoce como "falta de memoria". Se dice que la distribución geométrica "no tiene memoria". Esta característica también la tiene su análoga continua, la distribución exponencial negativa. ¿De qué se trata? La distribución geométrica no es afectada por lo que vino antes. Es decir, no importa desde cuándo empezamos a contar, siempre la probabilidad de las distintas cantidades de intentos hasta alcanzar un éxito estará distribuida de la misma forma. No importa si empezamos a contar justo después de un éxito, o después de una racha de 30 fracasos. Consideremos por ejemplo que en una determinada ciudad con muy mal clima, cada día tiene una probabilidad 0,2 de ser soleado. Y nuestro problema consiste en ver cuántos días tendremos que esperar para ver un día soleado. El siguiente calendario muestra los valores que resultan salir:

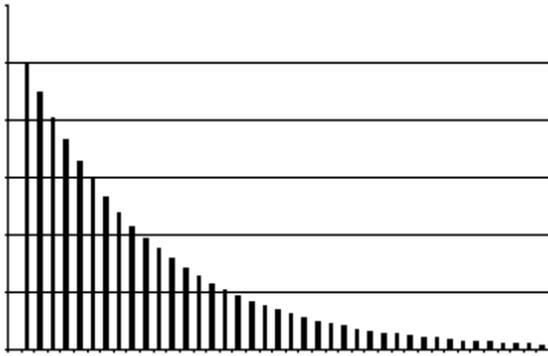
DOM	LUN	MAR	MIE	JUE	VIE	SAB
		1 	2 	3 	4 	5 
6 	7 	8 	9 	10 	11 	12 
13 	14 	15 	16 	17 	18 	19 

Como la distribución geométrica no tiene memoria, la cantidad de días que íbamos a tener que esperar para tener un día soleado estaba distribuida exactamente igual tanto después del martes 8 (que veníamos de una racha de 3 días malos) como después del martes 15 (que acabábamos de tener 2 días buenos seguidos) como antes del martes 1 (que ni siquiera sabíamos lo que había venido antes). Dicho de otro modo, si hoy es el martes 8 a la noche, y nos preguntan cuál es la probabilidad de que haya que esperar 3 días hasta el próximo día soleado, esa probabilidad es exactamente la misma que la que responderíamos si hoy fuera el martes 15 a la noche o martes 1 a la noche. Entonces sin importar en la noche de qué día nos paremos, siempre la cantidad de días que tendremos que esperar hasta que haya un día soleado está distribuida **exactamente igual** porque la distribución geométrica **no recuerda** lo que vino antes.

En la distribución binomial, la X tenía probabilidad no nula para un conjunto finito de valores, comprendidos entre 0 y n inclusive. En cambio la distribución geométrica tiene probabilidad no nula para infinitos valores, porque por ejemplo no es **imposible** tener que repetir el experimento 40 veces para conseguir un éxito. Es decir que no hay un x **máximo** para el cual $P(X = x)$ no es nulo (aunque de todos modos, para x suficientemente grande, $P(X = x)$ resultará despreciable).

Recordemos que para poder utilizar el modelo geométrico necesitamos suponer que todos los intentos de lograr el objetivo son independientes entre sí.

Aspecto



Vemos que cualquier valor a partir del 1 tiene probabilidad no nula. El 1 siempre es el valor más probable, y luego la probabilidad va descendiendo asintóticamente hacia el cero, pero nunca se hace cero debido a que no es imposible que el primer éxito se alcance en el intento 8.385.943

Problemas típicos

1) Necesitamos establecer una conexión. Cada vez que intentamos conectarnos, tenemos una probabilidad de 0,2 de lograr establecer la conexión.

- a) ¿Cuál es la probabilidad de que logremos conectarnos en menos de 4 intentos?**
- b) ¿Cuántas veces es de esperar que tengamos que intentar conectarnos hasta lograrlo?**
- c) Si cada intento nos lleva 20 segundos y además perdemos 10 segundos entre intento e intento para dejar todo listo para volver a intentar, ¿cuánto tiempo se espera que nos lleve el proceso de conectarnos?**

Resolución:

a) La mínima cantidad de intentos va a ser 1. Menos de 4 intentos significa "hasta 3 intentos inclusive". Es decir que lograr la conexión "en menos de 4 intentos" significa lograrla en el primer intento o en el segundo o en el tercero.

Tomamos $X: \text{Geom}(p = 0,2)$.

$$\Rightarrow P(\text{lograr la conexión en menos de 4 intentos}) = P(X=1) + P(X=2) + P(X=3) = p \cdot (1-p)^0 + p \cdot (1-p)^1 + p \cdot (1-p)^2 = 0,488$$

Otros ejercicios de este tipo, a lo sumo tendrán la dificultad de que el parámetro p no sea dato sino que haya que conseguirlo de diversas otras maneras, como se ve en el ejemplo 3 de la binomial.

$$b) E_x = \frac{1}{p} = 5$$

c) Teníamos:

X: la cantidad de intentos que nos lleva conectarnos y tomamos:

T: el tiempo que nos lleva el proceso

Entonces podemos poner:

$$T = 20 \cdot X + 10 \cdot (X - 1) = 20 \cdot X + 10 \cdot X - 10 = 30 \cdot X - 10$$

Como la esperanza es un operador lineal, hacemos:

$$E[T] = E[30 \cdot X - 10] = 30 \cdot E[X] - 10 = 140 \text{ segundos}$$

2) El 50% de los disparos da en el blanco. ¿Cuál es el mínimo de disparos que necesitaremos para tener 90% de confianza de dar en el blanco?

Resolución:

Si por ejemplo dijéramos que el mínimo es 5 disparos, no significa que haciendo 5 disparos se obtendrán fallos en las 4 primeras veces y éxito en la quinta. Tampoco significa siquiera que habrá 4 fallos y 1 éxito. Si nuestra respuesta fuera 5, estaríamos diciendo que hay probabilidad 0,9 de que el primer éxito se logre en uno de los primeros 5 disparos. Si el primer éxito se encontrara, por ejemplo, en el 3er disparo, no nos importa si luego se logran o no éxitos en el 4to y 5to disparo, logrando 2 ó incluso 3 éxitos. Sólo nos importa que el primer éxito se encuentre entre los primeros 5 intentos, porque esa es la única condición que tenemos que pedir para dar en el blanco en 5 o menos intentos.

Este problema no es como el anterior, porque en vez de preguntarnos la probabilidad, nos están dando la probabilidad y nos están preguntando cuál debe ser la condición sobre la variable para encontrar ese valor. En este caso la condición es " $X \leq m$ " y el problema consiste en buscar el m para satisfacer la probabilidad que nos dan.

Planteamos:

X:Geom(p = 0,5)

Queremos hallar m tal que:

$$P(X \leq m) \geq 0,9$$

Con lo cual el problema se reduce a sumar $P(X = 1) + \dots + P(X = m)$ hasta alcanzar 0,9. Es decir:

$$\sum_{i=1}^m P(X = i) \geq 0,9$$

Usando la fórmula de la distribución geométrica obtenemos:

$$P(X = 1) = 0,50000$$

$$P(X = 2) = 0,25000 \quad \Rightarrow \quad P(X \leq 2) = 0,75000$$

$$P(X = 3) = 0,12500 \quad \Rightarrow \quad P(X \leq 3) = 0,87500$$

$$P(X = 4) = 0,06250 \quad \Rightarrow \quad P(X \leq 4) = 0,93750$$

Con lo cual diremos que efectuando 4 disparos, tendremos más del 90% de confianza de acertar al blanco.

3) Juan y Pedro salen a cazar patos. Cada uno se va por su cuenta, y vuelve habiendo cazado un pato. La probabilidad de acertar un disparo es de 0,2. ¿Cuál es la probabilidad de que entre los 2 hayan hecho exactamente 8 disparos?

Resolución:

Este ejemplo ilustra las propiedades reproductivas de la distribución geométrica.

Tomamos:

X: cantidad de disparos hechos por Juan.

Y: cantidad de disparos hechos por Pedro.

Con lo cual:

X:Geom($p = 0,2$)

Y:Geom($p = 0,2$)

Y queremos obtener:

$$Z = X + Y$$

Suponiendo que los dos amigos son estadísticamente independientes, tenemos que:

Z:Pas($k = 2 ; p = 0,2$)

Luego usamos la fórmula de Pascal (se da en la siguiente sección) y obtenemos:

$$P(Z = 8) = 0,0734$$

4) En el acoplamiento de una estación espacial, el 20% de los intentos es exitoso. Calcule la probabilidad de que:

a) se logre el acoplamiento en 3 ó menos intentos

b) se logre el acoplamiento en 10 o menos intentos, sabiendo que se falló en los primeros 7.

c) ¿qué conclusión puede sacar de los resultados obtenidos en a y b?

Resolución:

a) Llamando X a la variable aleatoria a la cantidad de intentos necesarios hasta lograr el acoplamiento, queda:

X:Geom($p = 0,2$)

Con lo cual:

$$P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 0,488$$

b) Bajo las mismas condiciones que teníamos en a:

$$P(X \leq 10 / X > 7) = \frac{P(X \leq 10 \wedge X > 7)}{P(X > 7)} = \frac{P(7 < X \leq 10)}{P(X > 7)} = \frac{\sum_{i=8}^{10} P(X=i)}{1 - \sum_{i=1}^7 P(X=i)} = \frac{0,10234}{0,20972} = 0,488$$

c) Observamos que la probabilidad de que se necesiten 3 ó menos intentos, sin saber qué había pasado antes, es igual a la probabilidad de que se necesiten 3 ó menos intentos más, sabiendo que acaba de haber 7 fracasos seguidos. Esto nos muestra que la distribución geométrica no tiene memoria, porque puedo pararme antes de cualquier intento, y la probabilidad de que la cantidad de intentos necesarios cumpla tal o cual condición a partir de ese momento es la misma, sin importar a partir de cuándo comencemos a contar.

Distribución de Pascal

"¿Cuál es la probabilidad de obtener el k-ésimo éxito en el intento número x?"

Si

X:Pas (k ; p)

es decir: X es una variable de pascal con parámetros k y p

es decir: X es la variable que representa el número del intento en el cual se obtiene el éxito número k en experimentos de Bernoulli independientes cada uno con probabilidad de éxito p

$$P(X = x) = \begin{cases} \binom{x-1}{k-1} p^k \cdot (1-p)^{x-k} & x \geq k \\ 0 & \forall \text{ otro } x \end{cases}$$

entonces:

$$E(X) = \frac{k}{p}$$

$$\sigma_x^2 = \frac{k(1-p)}{p^2}$$

k es un número natural

p es un número real entre 0 y 1

Propiedades reproductivas

Si tenemos

- m variables X_i
- $X_i: \text{Pas}(k_i, p)$
- X_i independiente de X_j para $i \neq j$

$$Y = \sum_{i=1}^m X_i$$

•

entonces:

- $Y: \text{Pas}(k_Y, p)$

$$k_Y = \sum_{i=1}^m k_i$$

•

Es decir, la suma de m variables de pascal independientes cada una con igual p y con su propio k resulta ser una variable de pascal con el mismo p que las anteriores y k dado por la suma de los k de las variables originales.

Estrategia

Sabemos que nos encontramos frente a una distribución de pascal cuando:

- nos describen un experimento de Bernoulli (probabilidad de que una determinada pieza sea defectuosa: 0,2; probabilidad de que una operación resulte exitosa 0,9; etc.)
- nos dicen que vamos a seguir hasta el k -ésimo éxito (hasta que encontremos 500 piezas no falladas; hasta lograr 8 operaciones exitosas; etc.)
- nos preguntan cuál es la probabilidad de que logremos el objetivo en menos de x repeticiones, o la probabilidad de que nos tome más de x intentos lograr el objetivo, o la probabilidad de que lo logremos exactamente en el x -ésimo intento.

Al igual que sucedía con la binomial, la principal dificultad con la distribución de Pascal, una vez reconocida, puede consistir en conseguir la probabilidad de que un intento resulte exitoso. Luego para averiguar la cantidad de intentos necesarios para obtener k éxitos el uso de la fórmula es bastante inmediato.

Existe un caso particular de la distribución de Pascal, denominado distribución geométrica. Dicha distribución es una Pascal en la cual $k = 1$. Por eso la distribución geométrica sólo tiene el parámetro p . Generalmente y a menos que el problema sea demasiado obvio, no conviene hablar de las distribuciones geométrica y de Pascal como cosas distintas. De hecho la suma de variables geométricas da una variable de Pascal. Y esto no es sorprendente, porque al sumar las variables de Pascal de igual p se obtiene otra variable de pascal con la suma de las k . Entonces la suma de 8 variables geométricas con un determinado p resulta ser una variable de Pascal con $k = 8$ (y con el mismo p que las geométricas). Esperar 8 veces a tener un éxito (8 geométricas) es como esperar, empezando de cero, hasta el 8vo éxito (Pascal con $k = 8$).

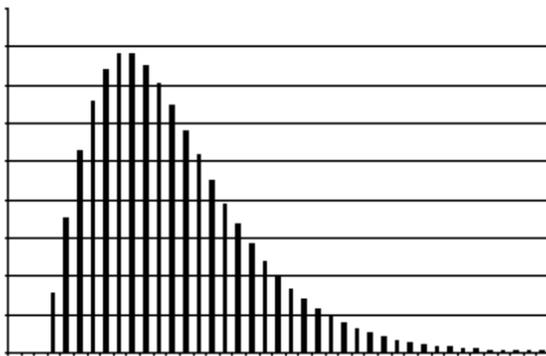
En la distribución binomial, la X tenía probabilidad no nula para un conjunto finito de valores, comprendidos entre 0 y n inclusive. En cambio la distribución de Pascal tiene probabilidad no nula para infinitos valores, porque por ejemplo no es **imposible** que el éxito número 28 se consiga en el intento 35.432.323. Es decir que

no hay un x **máximo** para el cual $P(X=x)$ no es nulo (aunque de todos modos, para x suficientemente grande, $P(X=x)$ resultará despreciable).

Pero **sí** hay un x **mínimo** para el cual la probabilidad es no nula, porque por ejemplo el éxito número 8 no puede ser obtenido en el intento número 5. Resulta importante recordar esto, especialmente cuando se trabaja con sumatorias que contienen probabilidades de pascal, para no cometer el error conceptual de incluir en la sumatoria términos en los cuales $x < k$ que en consecuencia serán nulos. Esto se ve más claramente en uno de los ejemplos.

Por último recordemos que para poder aplicar la distribución de Pascal es necesario que todos los intentos sean independientes.

Aspecto



El cero, y todos los valores menores que k , tienen probabilidad nula, debido a que k es la cantidad mínima de intentos para lograr k éxitos. A partir de k , la probabilidad crece con mayor o menor velocidad dependiendo de p , y luego de llegar al valor más probable, decrece lenta y asintóticamente hacia el 0.

Problemas típicos

1) Arrojo un dado hasta que obtengo por cuarta vez un 2. ¿Cuál es la probabilidad de que lo haya arrojado 10 veces?

Resolución:

Suponiendo que todas las veces que arrojé el dado son independientes, y que el dado es honesto, y que la distribución de probabilidad de sus caras se mantiene constante, puedo tomar:

$X: \text{Pas}(k = 4 ; p = 1/6)$

$$P(X = 10) = \text{comb}(x-1, k-1) \cdot p^k \cdot (1-p)^{x-k} = \text{comb}(9, 3) \cdot (1/6)^4 \cdot (5/6)^6 = 0,02171$$

2) En cada transmisión se envía un paquete de información. El 90% de los paquetes se transmite correctamente. Se necesita enviar 10 paquetes. Si un paquete no se transmitió correctamente, se repite la transmisión hasta que se reciba correctamente. Calcule la probabilidad de:

- a) Emplear 8 transmisiones para completar el trabajo.**
- b) Emplear menos de 13 transmisiones para completar el trabajo.**
- c) Si cada transmisión toma 20 μ s, y se pierden 10 μ s luego de cada transmisión preparando la siguiente, ¿cuánto tiempo es de esperar que tome completar el trabajo?**

Resolución:

a) Es imposible usar menos de k transmisiones para lograr k éxitos $\Rightarrow P = 0$

b) Tomamos:

X: cantidad de transmisiones necesarias para enviar los 10 paquetes.

Con lo cual:

X:Pas(k = 10 ; p = 0,9)

$$P(X < 13) = \sum_{i=0}^{12} P(X = i)$$

Pero son necesarias al menos k transmisiones, con lo cual

$$\sum_{i=0}^{12} P(X = i) = \sum_{i=10}^{12} P(X = i) = P(X = 10) + P(X = 11) + P(X = 12) = 0,88913$$

c) Teníamos:

X: la cantidad de transmisiones necesarias para completar el trabajo.

y tomamos:

T: el tiempo que nos lleva completar el trabajo

Entonces podemos poner:

$$T = 20 \cdot X + 10 \cdot (X - 1) = 20 \cdot X + 10 \cdot X - 10 = 30 \cdot X - 10$$

Como la esperanza es un operador lineal, hacemos:

$$E[T] = E[30 \cdot X - 10] = 30 \cdot E[X] - 10$$

Como X es una variable de pascal, $E_x = k/p = 11,111$

Entonces:

$$E[T] = 30 \cdot E_x - 10 = 323,33 \mu s$$

3) Juan y Pedro revisan cada uno una bolsa de tornillos surtidos. El 10% de los tornillos sirven. Juan necesita 6 tornillos, y Pedro necesita 8. ¿Cuántos tornillos estima Ud. que revisarán entre los dos hasta encontrar cada uno lo que necesita?

Resolución:

Este ejemplo ilustra las propiedades reproductivas de la distribución de Pascal.

Vemos que la cantidad de tornillos que revisará Juan hasta que encuentre lo que

necesita es una variable de pascal X:Pas(p = 0,1 ; k = 6). Y la cantidad que revisará

Pedro hasta que encuentre lo que necesita es una variable de pascal $Y: \text{Pas}(p = 0,1 ; k = 8)$. Visto que las p son iguales, y considerando a X e Y independientes, podemos establecer que si la cantidad de tornillos que revisarán entre los dos es $Z = X + Y$ entonces sabremos que $Z: \text{Pas}(p = 0,1 ; k = 14)$. Luego $E_Z = k / p = 140$.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 12 de julio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 6 de junio de 2004.

CAPÍTULO V

Proceso de Poisson

Es un proceso que consiste en considerar un continuo, en el cual ocurren eventos. Si por ejemplo consideramos la cantidad de fallas que una máquina tiene en 3 horas, el continuo es el tiempo, y los eventos son las fallas de la máquina. Otro ejemplo puede ser considerar la cantidad de muertes por determinada enfermedad en un año. Pero el continuo al que nos referimos no tiene necesariamente que ser tiempo. Por ejemplo podemos considerar un rollo de tela de 100 metros de longitud y contar la cantidad de manchas en ese tramo. En ese ejemplo, el continuo es la tela y los eventos las manchas.

Se definen las siguientes variables:

- T : la longitud de un intervalo del continuo que va a estudiarse.
- k : la cantidad de eventos que hay en ese intervalo.
- λ : la cantidad esperada de eventos por unidad de tiempo (intensidad).

Ejemplo

Si una máquina falla habitualmente en promedio 2 veces por hora, y la controlamos durante determinadas 3 horas y falla 7 veces, tenemos:

- $T = 3$ horas
- $k = 7$ eventos
- $\lambda = 2$ eventos / hora

Generalmente conocemos el valor de λ , y entonces nos preguntamos cuántos eventos obtendremos en una determinada cantidad de tiempo, o cuánto tiempo tendremos que esperar hasta observar una determinada cantidad de eventos.

De esta forma obtenemos 2 distribuciones:

- Poisson: consiste en preguntar por la cantidad de eventos en el período T . Es decir, dado T , calcular la distribución de k .
- Gamma: consiste en preguntar por la cantidad de tiempo necesario hasta observar k eventos. Es decir, dado k , calcular la distribución de T .

Y además:

- Exponencial negativa: caso particular de Gamma cuando $k = 1$, es decir, consiste en preguntar por la cantidad de tiempo necesaria hasta obtener el primer evento.

Distribución de Poisson

"¿Cuál es la probabilidad de obtener x eventos en el intervalo estudiado?"

Si bien el proceso de Poisson trabaja con los parámetros T (longitud del intervalo) y λ (intensidad), la distribución de Poisson usa solamente el parámetro $\mu = \lambda \cdot T$. Como T es la longitud del intervalo, y λ es la cantidad esperada de eventos por unidad de tiempo, entonces μ resulta ser la media. Es decir que esta distribución tiene la característica de que su media resulta valer directamente lo mismo que valga el parámetro μ .

Si

X:Pois(μ)

es decir: X es una variable Poisson con media μ .

es decir: X es la variable que representa la cantidad de eventos obtenidos en un intervalo de longitud T e intensidad λ .

entonces:

$$P(X = x) = \begin{cases} \frac{e^{-\mu} \mu^x}{x!} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$E(X) = \mu$$

$$\sigma_x^2 = \mu$$

μ es un número real positivo

Propiedades reproductivas

Si tenemos

- m variables X_i
- X_i :Pois(μ_i)
- X_i independiente de X_j para $i \neq j$

$$Y = \sum_{i=1}^m X_i$$

-

entonces:

- Y :Pois(μ_y)

$$\bullet \quad \mu_y = \sum_{i=1}^m \mu_i$$

Es decir, la suma de m variables Poisson independientes cada una con su propio μ resulta ser una variable Poisson con μ dado por la suma de los μ de las variables originales.

Estrategia

Sabemos que nos encontramos frente a la necesidad de emplear una distribución Poisson cuando existe un determinado intervalo en el cual suceden eventos, y necesitamos calcular cuántos eventos sucederán en dicho intervalo.

Puede ser que nos den la longitud del intervalo y la intensidad, o que directamente nos den la media.

- Cuando nos dan la longitud del intervalo y la intensidad:
 - T . El intervalo es continuo, pero no tiene por qué necesariamente ser tiempo. Ejemplos de intervalos: 2 horas, 3 metros de tela, 10 km. de una ruta, etc. Siempre será un número multiplicado por una unidad de medida, o algo que deba ser interpretado o tomado como una unidad de medida.
 - λ . La intensidad es la cantidad esperada de eventos por unidad de tiempo. Ejemplos de intensidades: 4 visitantes por hora, 5 fallas por metro de tela, 3 baches por km., etc. Vemos que siempre sus unidades serán una unidad de evento (visitantes, fallas, baches, etc.) dividida por una unidad de medida del mismo tipo que la del intervalo (Es decir, si el intervalo es 3 metros de tela, es decir, longitud de tela, la intensidad deberá ser una cierta cantidad de algo por unidad de longitud de la tela, por ejemplo 5 fallas por metro de tela). Si nos dieran la intensidad al revés (Ej.: en vez de 3 baches/km. , $1/3 \text{ km./bache}$) sólo hay que acomodarla haciendo 1 sobre eso.
 - Luego podremos obtener la media como $\mu = \lambda \cdot T$. La media quedará del estilo 8 visitantes, 15 fallas, 30 baches, etc. Siempre su unidad será la misma que la unidad de evento que aparecía en el numerador de la intensidad.
- Cuando nos dan directamente la media: puede ser que directamente nos digan el valor de la media, o que nos digan, por ejemplo, "3 errores por página", en un contexto donde se sobreentiende que estamos hablando de una (y sólo una) página.

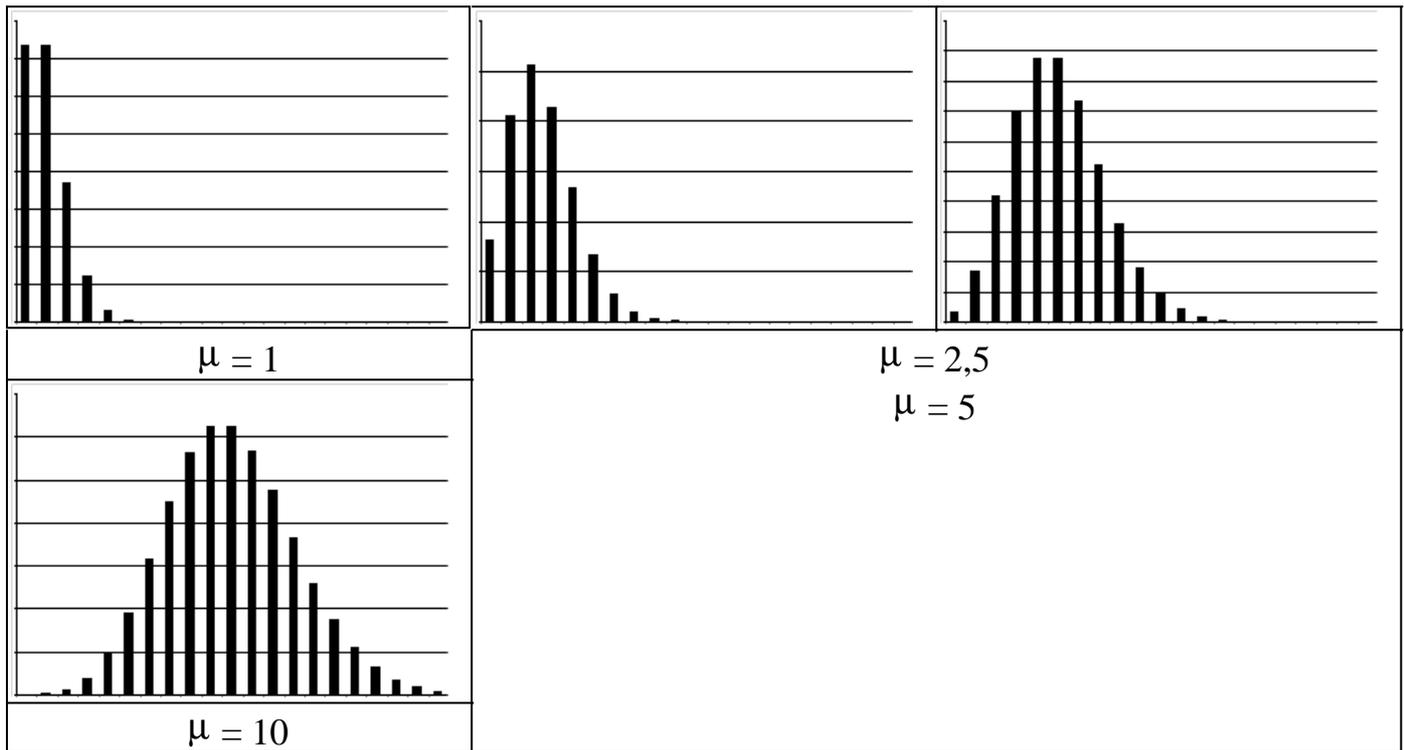
Notemos que la en la media ya están "incluidos" tanto la intensidad como la duración, y por lo tanto una distribución con 2 eventos/hora en 5 horas, será idéntica a una distribución con 1 evento/hora en 10 horas.

Una vez determinada la media, el problema ya no tiene mucha dificultad. No debemos olvidar suponer que el hecho de que en un determinado momento ocurra un evento, no nos afecta la probabilidad de tener o no más eventos, y cuándo

ocurrirán.

Si X es una variable de Poisson, $P(X = x)$ tendrá valor no nulo $\forall x \in [0 ; +\infty]$. Para $x < 0$, la probabilidad es nula. Vemos que los valores con probabilidad no nula están acotados inferiormente, pero no superiormente. No es imposible que en 2 horas halla 4039483 fallas.

Aspecto



Problemas típicos

1) A un comercio llegan en promedio 12 clientes por hora. El dueño debe salir durante 15 minutos a hacer una diligencia.

- ¿Cuál es la probabilidad de que no pierda ningún cliente?
- ¿Cuál es la probabilidad de que pierda 4 clientes?
- ¿Cuál es la probabilidad de que pierda 2 ó más clientes?

Resolución:

a) Debemos comenzar por advertir que tenemos un continuo (el tiempo) en el cual van a llegar clientes (eventos). Y además conocemos tanto la longitud del intervalo (15 minutos) como la cantidad esperada de eventos por unidad de tiempo ($^{12 \text{ clientes}}/\text{hora}$).

Entonces la cantidad de clientes que van a llegar en el intervalo estudiado (los 15

minutos en que el dueño no está) está dada por una distribución Poisson con media $\mu = \lambda \cdot T$.

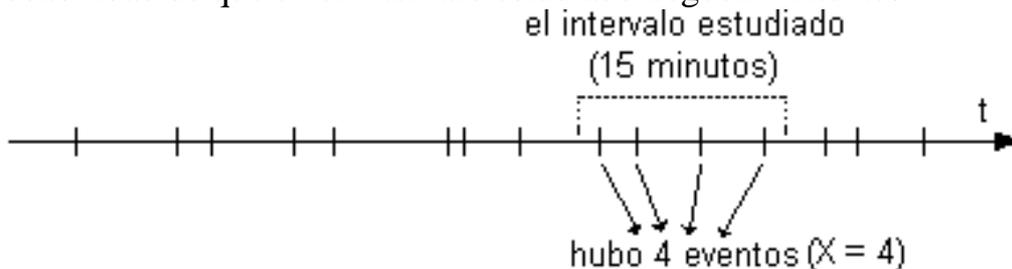
$$\mu = \lambda \cdot T = \frac{12 \text{ clientes}}{\text{hora}} \cdot \frac{15 \text{ minutos}}{60 \text{ minutos}} = \frac{12 \text{ clientes}}{60 \text{ minutos}} \cdot \frac{15 \text{ minutos}}{60 \text{ minutos}} = 3 \text{ clientes}$$

Luego definimos la variable aleatoria X : $\text{Pois}(\mu = 3)$. Notemos que la "unidad" del μ es "clientes", es decir, "eventos", y que esto ya tiene incluidos la longitud del intervalo y la intensidad. Entonces por ejemplo una variable Poisson con período 2 horas e intensidad 3 eventos por hora, está distribuida exactamente igual que una variable Poisson con período 1 hora e intensidad 6 eventos por hora.

$$P(X = x) = \frac{e^{-\mu} \cdot \mu^x}{x!} \Rightarrow P(X = 0) = \frac{e^{-3} \cdot 3^0}{0!} = 0,04979$$

Nos piden la probabilidad de no perder ningún cliente, es decir, la probabilidad de que en el intervalo estudiado no llegue ningún cliente, es decir, la probabilidad de que $X = 0$.

b) Bajo las mismas condiciones del problema anterior, ahora nos preguntan la probabilidad de que en el intervalo estudiado lleguen 4 clientes.



$$P(X = 4) = \frac{e^{-3} \cdot 3^4}{4!} = 0,16803$$

c) Bajo las mismas condiciones del problema anterior, ahora nos preguntan la probabilidad de que en el intervalo estudiado lleguen 2 ó más clientes.

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 0,80085$$

2) Una determinada máquina necesita ser reparada, en promedio, 8 veces por día. ¿Cuál es la probabilidad de que tenga que ser reparada menos de 3 veces?

Resolución:

Tenemos que comenzar por ver que no nos dan el intervalo en forma explícita. Tendremos que asumir que se refieren a un día.

Además hay otra suposición que también hacemos: durante el tiempo en que la máquina está siendo reparada, no está funcionando. Por lo tanto no se puede romper mientras se la está reparando. Por lo tanto mientras esté siendo reparada no hay eventos. Y entonces el hecho de que haya habido un evento nos va a afectar la

probabilidad de otros eventos, porque durante un tiempo después de él sabemos que no podrá haber eventos (porque la máquina estará siendo reparada y por lo tanto estará detenida y no habrá nuevas roturas). Para salvar esta situación, vamos a suponer que el tiempo que se tarda en reparar la máquina es muy corto y lo vamos a considerar despreciable.

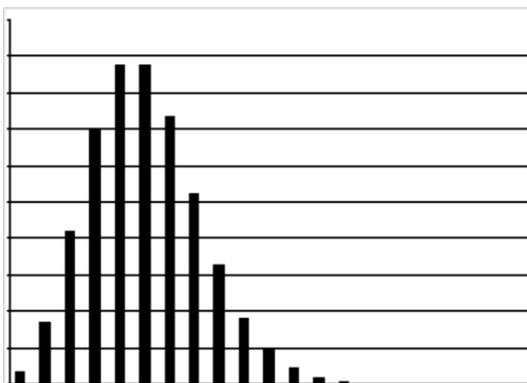
$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = 0,01375$$

3) En una ciudad hay en promedio 5 tormentas por año.

- a) ¿Cuál es la cantidad esperada de tormentas en un determinado año?
- b) ¿Cuál es la varianza?
- c) ¿Cuál es la cantidad más probable de tormentas en un determinado año?

Resolución:

- a) En la distribución de Poisson, la media coincide con el parámetro μ . La cantidad esperada de tormentas en un año cualquiera es 5.
- b) En la distribución de Poisson, la varianza también coincide con la media y el parámetro m . Por lo tanto también es 5.
- c) Aquí es necesario diferenciar entre media y valor más probable. Para una distribución discreta como la de Poisson, la media no es otra cosa que el promedio ponderado por la probabilidad entre todos los valores posibles. En cambio el valor más probable es el valor x tal que $P(X = x)$ es el máximo posible. Es decir, es el valor que más probabilidad tiene de ocurrir. Por ejemplo en un dado, la media es $1 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3,5$. La media es 3,5 aunque es imposible que obtengamos 3,5 al arrojar un dado. En cambio entre todos los valores posibles que pueden salir en un dado, la probabilidad más alta que algún valor tiene es $\frac{1}{6}$. Por eso los 6 números del dado son todos "el valor más probable". Volviendo al problema, una primera aproximación sería calcular $P(X = x)$ para varios valores de X , con lo cual obtendríamos algo así:



Veríamos que los valores de x que maximizan $P(X = x)$ son el 4 y el 5. Entonces los valores más probables son el 4 y el 5.

De hecho, en la distribución de Poisson se verifica que si el parámetro μ es entero,

entonces los valores más probables son siempre 2:
 μ ; $\mu-1$.

4) Un circuito falla, en promedio, 2 veces por hora.

- a) Calcular cuál es el tiempo que podrá funcionar tal que la probabilidad de que no falle sea de 0,88.**
- b) Responder a con una cantidad entera de minutos.**

Resolución:

a) Si X es una variable distribuida según Poisson que consista en la cantidad de fallas que tiene el circuito en un determinado período, entonces estamos buscando:
 $P(X = 0)$

Si la variable está distribuida según Poisson, tendrá su parámetro μ . Si nos preguntaran cuál es la media tal que la probabilidad de que no falle sea de 0,9 haríamos:

$$P(X = 0) = \frac{e^{-\mu} \mu^0}{0!} = 0,88 \Rightarrow e^{-\mu} = 0,88 \Rightarrow -\mu = \ln(0,88) \Rightarrow \mu = 0,12783$$

Pero nos están preguntando acerca de la cantidad de tiempo, es decir de T. Sabemos que $\mu = \lambda \cdot T \Rightarrow T = \mu/\lambda$.

λ era dato (2 fallas/hora) y μ lo acabamos de calcular (0,12783 fallas).

$$T = \frac{\mu}{\lambda} = \frac{0,12783 \text{ fallas}}{2 \text{ fallas}} \cdot \frac{\text{hora}}{1} = 0,063917 \text{ horas}$$

Esto nos dice que haciendo al circuito andar durante 0,063917 horas, hay probabilidad 0,88 de que no haya ninguna falla.

b) El problema es el mismo que en a, pero nos están pidiendo algo sobre el resultado: una cantidad ENTERA de minutos. Tendremos que transformar un valor no entero en algo entero. Como el tiempo es continuo, no hay problema en indicar algo como 0,063917 horas. Pero en cosas de naturaleza entera, por ejemplo la cantidad de intentos, no podemos informar como resultado un valor no entero. Es entonces cuando nos encontramos frente al problema de redondear.

Veamos:

$$\begin{array}{l} 1 \text{ hora} \quad \underline{\hspace{2cm}} \quad 60 \text{ minutos} \\ 0,063917 \text{ horas} \quad \underline{\hspace{2cm}} \quad 60 \cdot 0,063917 \text{ minutos} = 3,835 \text{ minutos} \end{array}$$

Tenemos que redondear. ¿Informaremos 3 ó 4 minutos?. Podríamos decir que como el número está más cerca del 4 que del 3, la respuesta es 4. Pero eso conlleva un error conceptual. Pensemos en la naturaleza del problema. 3,835 minutos garantizan que la probabilidad de que no haya fallas es 0,88. Si tomamos más tiempo, la probabilidad de que no haya fallas es menor. Y nos pidieron 0,88. No podemos dar una respuesta que nos de una probabilidad menor que la que nos

pidieron. Entonces la respuesta tiene que ser 3.

Si tomáramos 4 minutos, la probabilidad de que en 4 minutos no haya fallas es menor que 0,88 y no podemos dar un resultado con una probabilidad menor que la que nos pidieron. Por lo tanto, aunque 3,835 esté más cerca del 4 que del 3, tenemos que responder 3.

5) Se tienen 3 gallinas. Una de ellas pone en promedio 2 huevos por día. Otra pone en promedio 3 huevos por día. La restante pone en promedio 4 huevos por día. ¿Cuál es la probabilidad de que en un determinado día se produzcan exactamente 10 huevos?

Resolución:

Este ejemplo ilustra las propiedades reproductivas de la distribución Poisson.

Vamos a asumir que las gallinas ponen huevos independientemente, es decir, que la cantidad de huevos que pone una gallina no influencia la cantidad de huevos que ponen las otras.

Tenemos:

$$X_1: \text{Pois}(\mu_1 = 2)$$

$$X_2: \text{Pois}(\mu_2 = 3)$$

$$X_3: \text{Pois}(\mu_3 = 4)$$

$$Y = X_1 + X_2 + X_3$$

Con lo cual $Y: \text{Pois}(\mu_y)$ Donde $\mu_y = \mu_1 + \mu_2 + \mu_3 = 9$. Luego :

$$P(X = 10) = \frac{e^{-9} 9^{10}}{10!} = 0,11858$$

Distribución Exponencial Negativa

"¿Cuál es la probabilidad de tener que esperar x tiempo hasta obtener el primer evento?"

Si

$X: \text{ExpNeg}(\lambda)$

es decir: X es una variable Exponencial Negativa con intensidad λ .

es decir: X es la variable que representa el tiempo que hay que esperar hasta obtener el primer evento en un continuo con intensidad de eventos λ .

entonces:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$E(X) = \frac{1}{\lambda}$$

$$\sigma_X^2 = \frac{1}{\lambda^2}$$

λ es un número real positivo

Propiedades reproductivas

Si tenemos

- m variables X_i
- $X_i: \text{ExpNeg}(\lambda)$
- X_i independiente de X_j para $i \neq j$

$$Y = \sum_{i=1}^m X_i$$

•

entonces:

- $Y: \text{Gamma}(k, \lambda)$
- $k = m$

Es decir, la suma de m variables exponenciales independientes, todas con igual λ , resulta ser una variable gamma con el mismo λ que las anteriores y k dado por la cantidad de variables exponenciales que estamos sumando, es decir, m .

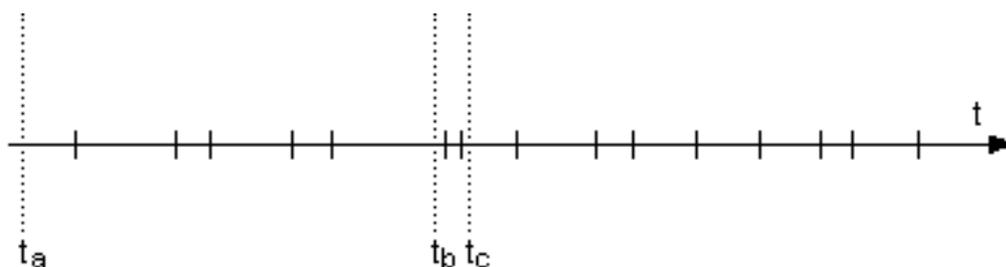
Estrategia

Sabemos que nos encontramos frente a una distribución exponencial negativa cuando:

- nos describen un continuo en el cual suceden eventos, como por ejemplo visitas a lo largo de un día, defectos a lo largo de una tela, fallas de un circuito a lo largo de un determinado período, etc.
- nos dicen que vamos a continuar observando hasta que suceda el primer evento. Ejemplo: hasta que llegue una visita, hasta que encontremos un defecto en la tela, hasta que el circuito falle, etc. Y también lo pueden decir al revés: mientras no llegue ninguna visita, mientras no encontremos un defecto en la tela, mientras el circuito no falle, etc.
- nos dan o nos permiten calcular la frecuencia promedio con que los eventos suceden (λ). Ejemplo: 3 visitas cada 15 minutos, 2 defectos por metro de tela, 3 fallas del circuito por día.
- nos preguntan acerca del tiempo, por ejemplo: cuál es la probabilidad de que el evento suceda en menos de x tiempo, la probabilidad de que tome más de x tiempo, etc.

Es importante saber que en un proceso Poisson, el intervalo de tiempo entre dos eventos consecutivos es siempre una variable exponencial negativa.

Otra característica de la distribución exponencial que es importante destacar, es lo que se conoce como "falta de memoria". Se dice que la distribución exponencial "no tiene memoria". Esta característica también la tiene su análoga discreta, la distribución geométrica. ¿De qué se trata? La distribución exponencial no es afectada por lo que vino antes. Es decir, no importa desde cuándo empezamos a contar, siempre la cantidad de tiempo que transcurrirá hasta que suceda el primer evento está distribuido de la misma forma. Dicho de otro modo, la probabilidad de que haya que esperar una determinada cantidad de tiempo hasta que haya un evento será la misma, tanto si empezamos a contar desde justo después de un evento como luego de una larga racha sin eventos. Por ejemplo veamos el siguiente continuo en el cual ocurren eventos:



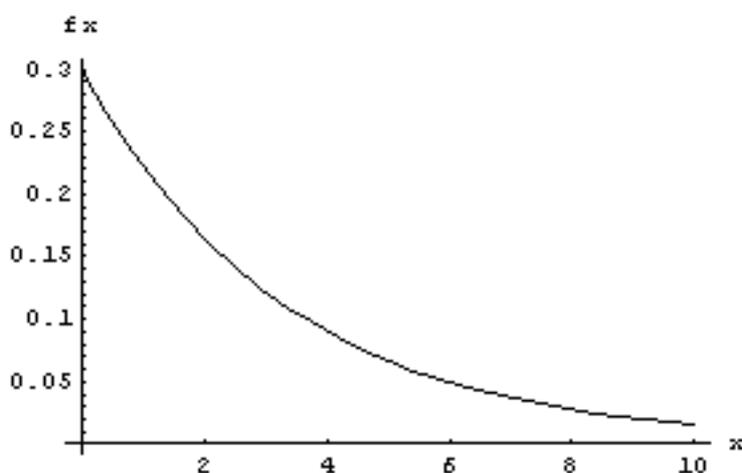
Si nos paramos en t_b y nos preguntamos cómo estará distribuido el tiempo que hay que esperar hasta tener un evento (luego de un gran período sin eventos), ese

tiempo estará distribuido probabilísticamente igual que el tiempo que habrá que esperar si estamos parados en t_c (donde acaban de suceder dos eventos prácticamente seguidos) y también estará distribuido igual que el tiempo que habrá que esperar si estamos parados en t_a (donde ni siquiera sabemos lo que pasó antes). Entonces sin importar dónde nos paremos, siempre la cantidad de tiempo que hay que esperar hasta el próximo evento está distribuida **exactamente igual** porque la distribución exponencial negativa **no recuerda** lo que vino antes.

La distribución exponencial negativa en realidad es un caso particular de la distribución gamma. Una variable exponencial puede ser vista como una variable gamma cuyo parámetro λ es el mismo que el de la exponencial, y cuyo parámetro k es igual a 1. De ahí que sumar variables exponenciales es en esencia como sumar variables gamma, y de ahí que la suma de variables exponenciales ES una variable gamma. Por esto, si sospechamos que en un problema tendremos que sumar variables exponenciales, puede resultar una idea bastante práctica considerarlas desde el principio variables gamma. De hecho la distribución exponencial se enseña separada de la gamma porque es más fácil aprender del caso particular al caso general.

La distribución exponencial es no nula para todos los tiempos mayores a cero, porque es imposible tener que esperar un tiempo negativo hasta el primer (o próximo) evento, pero no es imposible tener que esperar cualquier tiempo arbitrario por más grande que éste sea. Pero también es cierto que la probabilidad de tener que esperar un tiempo muy grande se hace despreciable.

Aspecto



Vemos que cualquier valor a partir del 0 tiene probabilidad no nula. El 0 será siempre el máximo de la función de densidad, y luego la probabilidad va descendiendo en forma asintótica hacia el cero, pero nunca se hace cero debido a que no es imposible que el primer evento ocurra en un tiempo arbitrariamente grande.

Problemas típicos

1) En promedio vienen 3 colectivos por hora, distribuidos según un proceso Poisson. ¿Cuál es la probabilidad de tener que esperar el colectivo más de 20 minutos?

Resolución:

Si la llegada de los colectivos puede ser vista como eventos en un continuo distribuidos según Poisson, entonces el tiempo de espera al llegar a la parada puede asociarse a una variable exponencial negativa. La intensidad del proceso es de 3 colectivos por hora, es decir, 3 eventos cada 60 minutos. Por lo tanto $\lambda = 0,05$. Si la variable exponencial negativa a la que hacíamos referencia, es decir, el tiempo de espera hasta el primer evento (que venga un colectivo) es X , entonces:

$X: \text{ExpNeg}(\lambda = 0,05)$

$$P(\text{esperar más de 20 minutos}) = P(X > 20) = \int_{20}^{+\infty} \lambda \cdot e^{-\lambda x} dx = 0,36787944$$

2) La duración de una lamparita está distribuida en forma exponencial negativa con una media de 300 horas. Calcule la probabilidad de que una lamparita:

- dure más de 100 horas.**
- dure más de 500 horas, sabiendo que duró más de 400 horas.**
- ¿qué conclusión puede sacar de los resultados obtenidos en a y b?**

Resolución:

a) Llamando X a la variable aleatoria asociada a la duración de la lamparita, si la media es 300 horas, entonces $E_X = 300 \Rightarrow \lambda = 1/E_X = 1/300$. Como X está distribuida en forma exponencial negativa, tenemos que $f_X = \lambda \cdot e^{-\lambda x}$ para $X > 0$, con lo cual:

$$P(X > 100) = \int_{100}^{+\infty} f_X \cdot dx = \int_{100}^{+\infty} \lambda \cdot e^{-\lambda x} \cdot dx = 0,71653131$$

b) Bajo las mismas condiciones que teníamos en a:

$$P(X > 500 / X > 400) = \frac{P(X > 500 \wedge X > 400)}{P(X > 400)} = \frac{P(X > 500)}{P(X > 400)} = \frac{\int_{500}^{+\infty} \lambda \cdot e^{-\lambda x} dx}{\int_{400}^{+\infty} \lambda \cdot e^{-\lambda x} dx} = 0,71653131$$

c) Observamos que la probabilidad de que dure más de 100 horas más, sin saber

qué había pasado antes es igual a la probabilidad de que dure más de 100 horas más, sabiendo que venía durando más de 400. Esto nos muestra que la distribución exponencial negativa no tiene memoria, porque puedo pararme en cualquier punto del continuo, y la probabilidad de que la duración de la lamparita cumpla tal o cual condición a partir de ese momento es la misma, sin importar a partir de qué punto comencemos a medir.

3) En una tela, las fallas se distribuyen según un proceso Poisson, a razón de 1 falla cada 15 metros. ¿Cuál es la probabilidad de que la distancia entre la 4^{ta} falla y la 5^{ta} falla sea mayor a un metro?

Resolución:

Este ejemplo muestra que, en un proceso Poisson, el intervalo entre dos eventos consecutivos es una variable exponencial negativa.

Entonces La distancia entre dos fallas consecutivas (sean éstas la 4^{ta} y la 5^{ta} u otras dos consecutivas cualesquiera) es una variable exponencial negativa con $\lambda = 1/15$.

$$P(X > 1) = \int_1^{+\infty} f_X \cdot dx = \int_1^{+\infty} \lambda \cdot e^{-\lambda x} \cdot dx = 0,9355$$

4) Se tiene un determinado sistema que funciona a baterías. La duración de cada batería es una variable aleatoria exponencial negativa, y su media es 10 horas. Cuando una batería se gasta, se reemplaza con otra. Si se tienen 4 baterías, calcule la probabilidad de que las baterías alcancen para menos de 40 horas.

Resolución:

Asumiendo que las baterías son independientes, tenemos 4 variables exponenciales negativas, independientes e idénticamente distribuidas. Como en la distribución exponencial negativa el parámetro intensidad λ es la inversa de la media, entonces λ , que supondremos igual para las 4 baterías, es 0,1. Tenemos:

$X_i: \text{ExpNeg}(\lambda = 0,1)$ con $i \in [1;4]$

$$Y = \sum_{i=1}^4 X_i$$

La suma de n variables exponenciales negativas independientes y con igual λ es una variable gamma con $k = n$ y el mismo λ que las exponenciales. Con lo cual:

$Y: \text{Gamma}(k = 4 ; \lambda = 0,1)$

Y nos piden:

$$P(Y < 40) = \int_0^{40} \frac{\lambda (\lambda x)^{k-1} e^{-\lambda x}}{k-1} dx$$

Con $k = 4$ y $\lambda = 0,1$. Usando la relación entre la distribución gamma y la distribución de Poisson descrita en la siguiente sección, podemos obtener el valor

de la integral tomando $Z: \text{Pois}(\mu)$, con $\mu = 40$ $\lambda = 4$, y haciendo:

$$\int_0^{40} \frac{\lambda (\lambda x)^{k-1} e^{-\lambda x}}{k-1} dx = 1 - \sum_{i=0}^{\infty} P(Z=i) = 1 - \sum_{i=0}^{\infty} P(Z=i) = 1 - \sum_{i=0}^{\infty} \frac{e^{-\mu} \mu^i}{i!} = 1 - \sum_{i=0}^{\infty} \frac{e^{-4} 4^i}{i!} = 0,56653$$

Distribución Gamma

"¿Cuál es la probabilidad de tener que esperar x tiempo hasta obtener el k -ésimo evento?"

Si

X:Gamma(λ ; k)

es decir: X es una variable Gamma con parámetros λ y k .

es decir: X es la variable que representa el tiempo que hay que esperar hasta obtener el evento número k , en un continuo con intensidad de eventos λ .

entonces:

$$f_X(x) = \begin{cases} \frac{\lambda(\lambda x)^{k-1} e^{-\lambda x}}{\Gamma(k)} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$E(X) = \frac{k}{\lambda}$$

$$\sigma_x^2 = \frac{k}{\lambda^2}$$

λ es un número real positivo

k es un número natural

$\Gamma(k)$ para k natural vale $(k-1)!$

$$\int_0^{x_0} f_X(x) dx = 1 - \sum_{i=0}^{x_0} P(Y = i)$$

Para facilitar el cálculo puede resultar útil:

donde X es la variable gamma con la que estamos trabajando, e Y es una variable de Poisson con $\mu = \lambda \cdot x_0$

Propiedades reproductivas

Si tenemos

- m variables X_i

- $X_i: \text{Gamma}(\lambda; k_i)$
- X_i independiente de X_j para $i \neq j$

$$Y = \sum_{i=1}^m X_i$$

-

entonces:

- $Y: \text{Gamma}(\lambda; k_y)$

$$k_y = \sum_{i=1}^m k_i$$

-

Es decir, la suma de m variables gamma independientes, todas con igual λ , resulta ser una variable gamma con el mismo λ que las anteriores y k dado por la suma de los k de las variables originales.

Cálculo

Como la función de densidad de la distribución gamma no es sencilla de integrar, se usa una forma alternativa de calcularla.

La probabilidad de que el tiempo que se tarda en obtener el k -ésimo evento sea menor que x_0 , es igual a la probabilidad de que en un intervalo de duración x_0 haya k ó más eventos (recomendamos dedicar un momento a comprender dicha afirmación).

Entonces $F_X(x_0) = 1 - F_Y(k-1)$, donde X es una variable gamma con parámetros λ y k , e Y es una variable de Poisson con parámetro $\mu = \lambda \cdot x_0$.

Reemplazando las F 's por sus definiciones, queda una expresión conveniente para calcular:

$$\int_0^{x_0} f_X(x) dx = 1 - \sum_{i=0}^{k-1} P(Y=i)$$

y también puede ser útil:

$$\int_{x_0}^{+\infty} f_X(x) dx = \sum_{i=0}^{k-1} P(Y=i)$$

De esta forma podemos calcular probabilidades de una distribución gamma recurriendo a las cuentas que se usan para la distribución Poisson, que son mucho más sencillas.

Estrategia

Sabemos que nos encontramos frente a una distribución gamma cuando:

- nos describen un continuo en el cual suceden eventos, como por ejemplo visitas a lo largo de un día, defectos a lo largo de una tela, fallas de un circuito a lo largo de un determinado período, etc.
- nos dicen que vamos a continuar observando hasta que suceda una determinada cantidad de eventos. Ejemplo: hasta que lleguen 5 visitas, hasta que encontremos 10 defectos en la tela, hasta que el circuito falle por 5ta vez, etc.
- nos dan o nos permiten calcular la frecuencia promedio con que los eventos suceden (λ). Ejemplo: 3 visitas cada 15 minutos, 2 defectos por metro de tela, 3 fallas del circuito por día.
- nos preguntan acerca del tiempo, por ejemplo: cuál es la probabilidad de que la cantidad de eventos indicada suceda en menos de x tiempo, la probabilidad de que tome más de x tiempo, etc.

En la función de densidad de la distribución gamma, aparece en el denominador la función Γ , "función gamma". Para números naturales, esta función se transforma en la función factorial, luego de restarle 1 al número. Ejemplo: $\Gamma(5) = 4!$. En el caso

$$\Gamma(k) = \int_0^{+\infty} x^{k-1} e^{-x} dx$$

general,

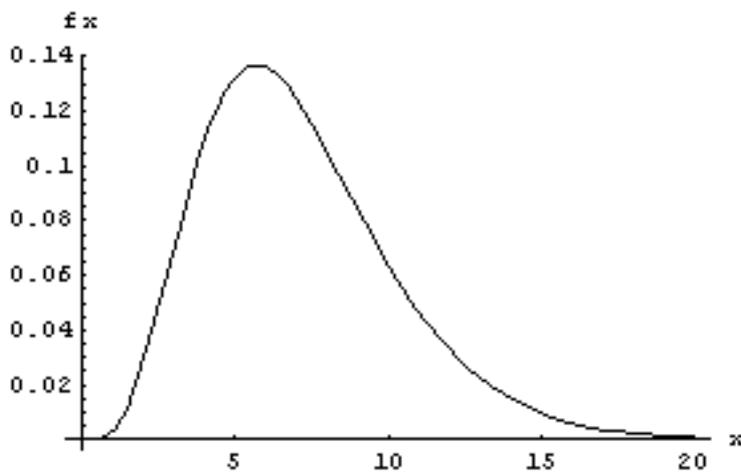
Pero como en la distribución gamma k siempre es natural, no utilizaremos dicha definición y adoptaremos $\Gamma(k) = (k-1)!$

Como vimos antes, la distribución gamma tiene un caso particular interesante: la distribución exponencial negativa. Dicha distribución es una gamma en la cual $k=1$. Por eso la distribución exponencial negativa sólo tiene el parámetro λ . Generalmente y a menos que el problema sea demasiado obvio, no conviene hablar de las distribuciones exponencial negativa y gamma como cosas distintas. De hecho la suma de variables exponenciales da una variable gamma. Y esto no es sorprendente, porque al sumar las variables gamma de igual λ se obtiene otra variable gamma con la suma de las λ . Entonces la suma de 8 variables exponenciales con un determinado λ resulta ser una variable gamma con $k = 8$ (y con el mismo λ que las exponenciales). Visto de otra forma, como en un proceso de Poisson el intervalo de tiempo entre dos eventos consecutivos está distribuido exponencialmente, entonces es natural que la distribución gamma (tiempo hasta k eventos) sea en esencia una suma de variables exponenciales negativas independientes. Esperar k veces hasta que ocurra un evento (k exponenciales) es lo mismo que esperar hasta el k-ésimo evento (gamma).

La distribución gamma es no nula para todos los tiempos mayores a cero, porque es imposible tener que esperar un tiempo negativo hasta que sucedan eventos, pero no es imposible tener que esperar cualquier tiempo arbitrario por más grande que éste sea. Pero también es cierto que la probabilidad de tener que esperar un tiempo muy grande se hace despreciable.

Por último recordemos que para poder aplicar la distribución gamma, todos los eventos deben ser independientes, como corresponde a los procesos de Poisson. Es decir, el hecho de que suceda un evento no aumenta ni disminuye la probabilidad de que haya más eventos en cualquier momento futuro.

Aspecto



Vemos que cualquier valor a partir del 0 tiene probabilidad no nula. La función de densidad crece hasta el máximo, y luego va descendiendo en forma asintótica hacia el cero, pero nunca se hace cero debido a que no es imposible que el k-ésimo evento ocurra en un tiempo arbitrariamente grande.

Problemas típicos

1) Un vendedor ambulante de loros tiene 5 loros para vender. En promedio se venden 2 loros por hora. ¿Cuál es la probabilidad de que le tome menos de 3 horas vender los 5 loros que tiene?

Resolución:

Si suponemos que la venta de cada loro es independiente (es decir que el hecho de que se haya vendido un loro no afecta ni la probabilidad de que se vendan o no más loros en el futuro, ni los momentos en que se vendan los otros loros) y asumimos que se trata de un proceso Poisson, entonces el tiempo que toma vender los 5 loros es una variable gamma con $\lambda = 2$ y $k = 5$. Es decir:

X:Gamma($\lambda = 2$; $k = 5$)

Luego, para los tiempos $x > 0$:

$$f_X(x) = \frac{\lambda (\lambda x)^{k-1} e^{-\lambda x}}{(k-1)!} = \frac{2(2x)^4 e^{-2x}}{24} = \frac{4}{3} x^4 e^{-2x}$$

Y luego la probabilidad de que se tarde menos de 3 horas es:

$$P(X < 3) = \int_{-\infty}^{+\infty} f_X(x) dx = \int_0^3 \frac{4}{3} x^4 e^{-2x} dx = 0,7149$$

Y esa es la respuesta. También se podría haber utilizado la relación con la variable de Poisson para obtener el resultado. Si X es una variable gamma con parámetros λ

$\int_0^{x_0} f_X(x) dx = 1 - \sum_{i=0}^{m-1} P(Y = i)$, donde Y es una variable de Poisson con $\mu = \lambda \cdot x_0$

En este caso x_0 vale 3, con lo cual $m = 6$, y la probabilidad pedida es:

$$1 - \sum_{i=0}^{m-1} P(Y = i) = 1 - \sum_{i=0}^{m-1} \frac{e^{-\mu} \mu^i}{i!} = 1 - \sum_{i=0}^{m-1} \frac{e^{-6} 6^i}{i!} = 1 - e^{-6} \sum_{i=0}^{m-1} \frac{6^i}{i!} = 0,7149$$

Y obtuvimos el mismo resultado que de la otra forma.

2) Un perro ladra según un proceso Poisson, a razón de 80 ladridos por hora. Un individuo se sienta junto al perro, pero al décimo ladrido se enoja y se retira. Otro día, se sienta junto al perro, y el cabo del ladrido número 15 se retira. Un tercer día, se retira luego del ladrido número 20. ¿Cuál es la probabilidad de que en total se pase más de media hora junto al perro?

Resolución:

El tiempo que el individuo pasa sentado junto al perro en un día cualquiera es:

X:Gamma($\lambda = 80$; k), donde k es la cantidad de ladridos que soporta ese día.

Luego:

X_1 :Gamma($\lambda = 80$; $k_1 = 10$)

X_2 :Gamma($\lambda = 80$; $k_2 = 15$)

X_3 :Gamma($\lambda = 80$; $k_3 = 20$)

Suponiendo que los 3 días son independientes entre sí, podemos decir que el tiempo que el individuo pasa sentado junto al perro en los 3 días es:

$$Z = X_1 + X_2 + X_3$$

Luego, por propiedades reproductivas de la distribución gamma, como Y es la suma de variables gamma con igual λ , queda:

Z:Gamma($\lambda = 80$; $k_y = 45$)

Por último, la probabilidad de que en total el individuo se pase más de media hora

junto al perro es $\int_{0,5}^{+\infty} f_Z(z) dz$, que se puede resolver integrando o bien mediante

$$\sum_{i=0}^{+\infty} P(Y = i)$$

Poisson con $i=0$. Como k es grande, conviene integrar, ya que usar Poisson implicaría sumar 45 términos. El resultado es 0,76568.

Observemos que el tiempo que toma esperar 10, 15 y 20 ladridos en 3 días distintos y el que lleva sentarse a esperar 45 ladridos está distribuido exactamente igual.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 6 de junio de 2004_

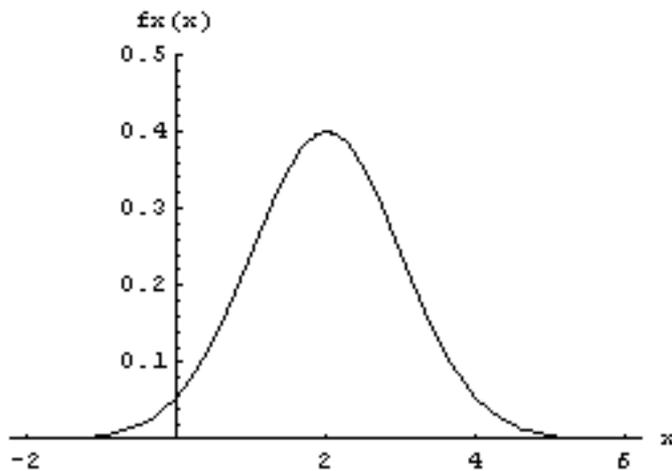
El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004_

Distribución Normal

Cuando la función de densidad es la siguiente:

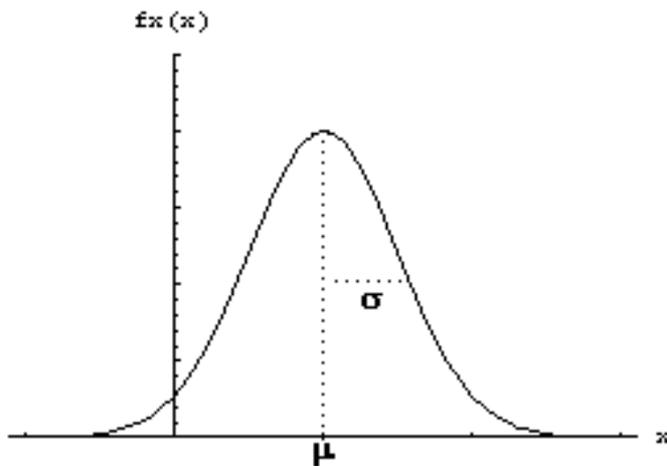


$$f_X(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi} \sigma} \quad \forall x \in \mathfrak{R}$$

la distribución se llama "Normal" (o de "Gauss").

La gráfica de esta función de densidad se conoce con el nombre de "campana de Gauss"

A primera vista podemos observar:

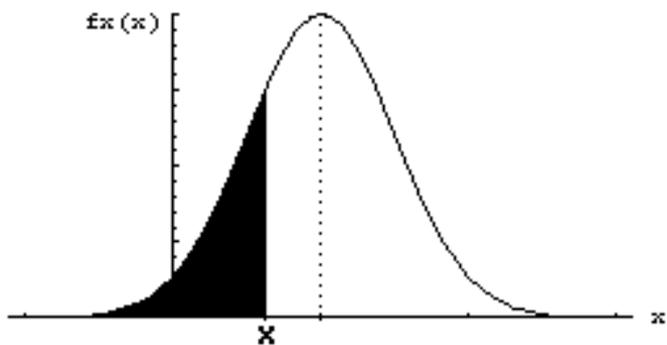


- a diferencia de todas las distribuciones que vimos anteriormente, es no-nula para todos los números reales.
- tiene 2 parámetros, μ y σ .

El parámetro μ puede ser cualquier número real, y es, directamente, la media de la distribución.

El parámetro σ puede ser cualquier número real positivo, y es, directamente, el desvío estándar de la distribución.

La notación $X \sim N(\mu; \sigma)$ significa que la variable aleatoria X tiene una distribución normal con parámetros μ y σ , o dicho de otra forma, que la variable aleatoria X tiene una distribución normal, cuya media es μ , y cuya varianza es σ^2 .



Como para todas las distribuciones continuas, para calcular probabilidades podemos plantear:

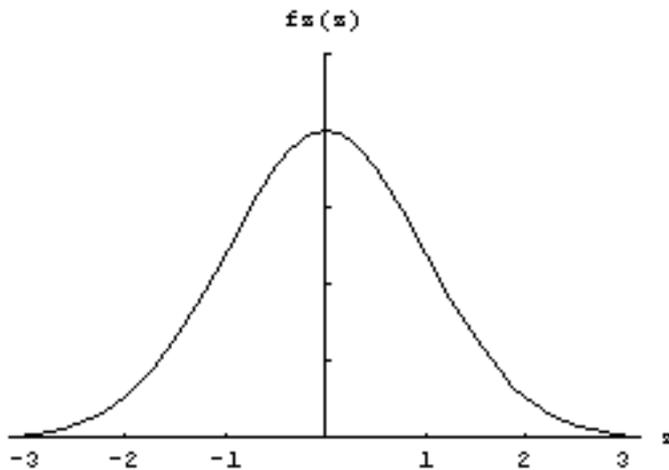
$$P(X \leq x) = F_X(x) = \int_{-\infty}^x f_X(x) dx$$

Sin embargo, a los fines prácticos, esta distribución presenta un problema: la integración de una función de la familia e^{x^2} no es un proceso simple. Por tal motivo,

en vez de integrar para encontrar el área bajo la curva, los valores de la función de distribución acumulada F se toman de una tabla (Ver apéndice D).

Observemos que, al ser μ y σ números reales, hay infinitas distribuciones posibles, y no se pueden tener infinitas tablas. Es por eso que se trabaja con una distribución particular denominada "normal estándar" y lo que se hace es transformar cualquier normal en una normal estándar, mediante un proceso denominado estandarización.

Distribución Normal Estándar



Cuando $\mu = 0$ y $\sigma = 1$, la distribución se llama normal estándar.

Se puede demostrar que si X es cualquier variable aleatoria normal, y tomamos la

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

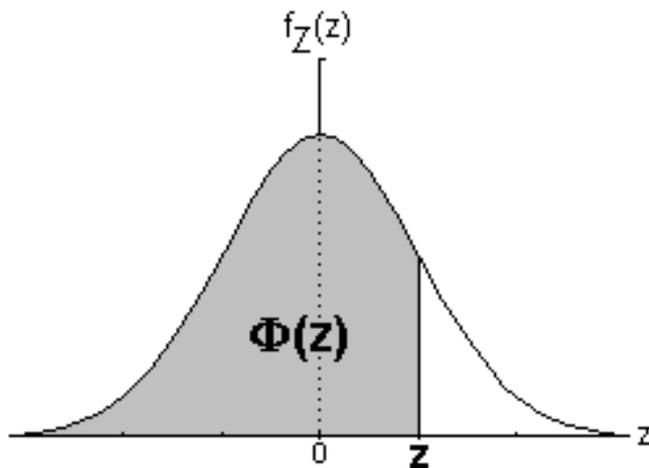
variable aleatoria Z , entonces Z resulta ser una variable aleatoria normal estándar.

Es decir:

$$X: N(\mu; \sigma) \wedge Z = \frac{X - \mu}{\sigma} \Rightarrow Z: N(0, 1).$$

lo cual puede ser demostrado mediante un simple cambio de variables.

Esto nos permite, dada cualquier variable aleatoria normal, encontrar una variable aleatoria normal estándar, que es la que encontraremos en las tablas. A la F_z la notaremos con la letra Φ . El proceso de tomar ese cambio de variables para obtener una normal estándar a partir de una normal se conoce con el nombre de estandarización.



Por ejemplo, si tenemos una variable aleatoria X y sabemos que sigue una distribución normal con parámetros μ y σ , y necesitamos calcular, $P(X \leq x)$, haremos:

$$P(X \leq x) = F_X(x) = F_Z\left(\frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

y el valor de Φ en ese punto lo tomamos de la tabla.

Ejemplo:

La longitud de los clavos fabricados por una máquina, en milímetros, es una variable aleatoria X que sigue una distribución normal, con media 10 y varianza 2. Calcular:

- 1) ¿Cuál es la probabilidad de que un clavo elegido al azar mida menos de 12 milímetros?
- 2) ¿Cuál es la probabilidad de que un clavo elegido al azar mida menos de 7 milímetros?

1) Tenemos: $X:N(10;2)$

Calculamos:

$$P(X \leq 12) = F_X(12) = F_Z\left(\frac{12 - 10}{2}\right) = F_Z(1) = \Phi(1)$$

De la tabla de la distribución normal estándar obtenemos que $\Phi(1) = 0,84134$. Entonces la probabilidad que estamos buscando es $P(X \leq 12) = 0,84134$

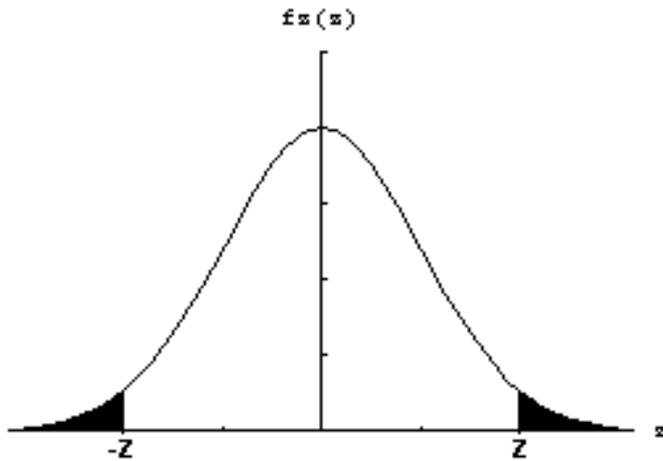
2) Análogamente hacemos:

$$P(X \leq 7) = F_X(7) = F_Z\left(\frac{7 - 10}{2}\right) = F_Z(-1.5) = \Phi(-1.5)$$

Y cuando vamos a buscar en la tabla $\Phi(-1.5)$ nos damos cuenta de que no se encuentra.

Puede suceder que la tabla que estemos usando comprenda solamente los valores

positivos de z . Es decir, que contenga solamente los valores de $\Phi(z)$ para $z > 0$. Tal es el caso de la tabla incluida en esta obra.



Si necesitamos calcular $\Phi(z)$ para algún $z < 0$, podemos valernos de la siguiente propiedad:

$$\Phi(-z) = 1 - \Phi(z)$$

En el gráfico podemos ver que, aunque lo que buscamos es el área sombreada de la izquierda, esta es igual al área sombreada de la derecha, la cual puede ser calculada usando un valor positivo de z (y que por lo tanto podremos encontrar en la tabla).

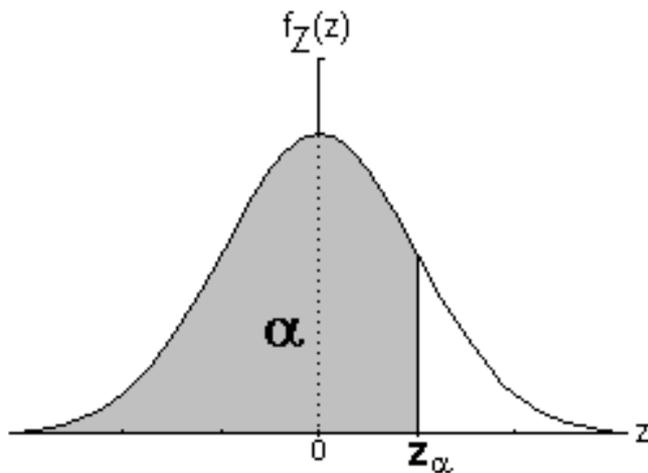
Fractiles

Ya sabemos cómo encontrar la probabilidad $P(X \leq x)$.

Pero el problema puede ser al revés: conociendo la probabilidad y la distribución, encontrar x tal que $P(X \leq x)$ sea dicha probabilidad. En otras palabras, encontrar x tal que el área acumulada a la izquierda de x sea igual a esa probabilidad. Ese valor de x se conoce como fractil.

Para una normal estándar, z_{α} quiere decir "el z a la izquierda del cual el área encerrada es α ".

Si por ejemplo tenemos que $P(X \leq x) = 0,95$ haremos:



$$P(X \leq x) = 0,95 \Rightarrow F_X(x) = 0,95 \Rightarrow$$

$$\Rightarrow \Phi\left(\frac{x - \mu}{\sigma}\right) = 0,95 \Rightarrow \frac{x - \mu}{\sigma} = z_{0,95}$$

De la tabla obtenemos que el z para que el área encerrada a la izquierda sea 0,95, es decir, $z_{0,95}$, es 1,645. Luego:

$$\frac{x - \mu}{\sigma} = 1,645 \Rightarrow x = 1,645\sigma + \mu$$

donde μ y σ son dato.

Ejemplo:

La longitud de los clavos fabricados por una máquina, en milímetros, es una variable aleatoria X que sigue una distribución normal, con media 10 y varianza 2. Se debe dar una especificación del máximo la longitud de los clavos, tal que el 90% de los clavos cumpla con la especificación. ¿Cuál debe ser la especificación?

Tenemos $X:N(10;2)$ y además nos piden que $P(X \leq x) = 0,9$

$$P(X \leq x) = 0,9 \Rightarrow F_X(x) = 0,9 \Rightarrow \Phi\left(\frac{x - 10}{2}\right) = 0,9 \Rightarrow \frac{x - 10}{2} = z_{0,9}$$

Usamos la tabla y obtenemos que $\Phi(1,28) = 0,9$

$$\frac{x - 10}{2} = 1,28 \Rightarrow x = 12,56$$

Con lo cual si decimos que la longitud máxima de los clavos debe ser de 12,56 el 90% de los clavos fabricados cumplirá con la especificación

Encontrar los parámetros

Otro problema posible es que sepamos que una variable aleatoria es normal pero no conozcamos los parámetros μ y σ . Si conociéramos, por ejemplo, para 2 valores x_1 y x_2 que la probabilidad de que X sea menor o igual a esos valores es p_1 y p_2 respectivamente, entonces podremos calcular el valor de los parámetros, es decir, la forma que la campana debe tener para que $P(X \leq x_1) = p_1$ y $P(X \leq x_2) = p_2$. Si

estandarizamos llegamos a que:

$$\Phi\left(\frac{x_1 - \mu}{\sigma}\right) = p_1 \quad \wedge \quad \Phi\left(\frac{x_2 - \mu}{\sigma}\right) = p_2$$

Conociendo p_1 y p_2 , de la tabla obtenemos z_{p_1} y z_{p_2} , con lo cual podemos plantear un sistema de 2 ecuaciones con 2 incógnitas, debido a que x_1 y x_2 también son dato.

$$\begin{cases} \frac{x_1 - \mu}{\sigma} = z_{p_1} \\ \frac{x_2 - \mu}{\sigma} = z_{p_2} \end{cases}$$

Y resolviendo el sistema conseguimos μ y σ .

Ejemplo:

La longitud de los clavos fabricados por una máquina, en milímetros, es una variable aleatoria X que sigue una distribución normal. Se sabe que el 80% de los clavos fabricados miden menos de 11mm, y que el 90% de los clavos fabricados miden menos de 12mm. ¿Cuál es la media y la varianza de los clavos producidos por la máquina?

Sabemos que $P(X \leq 11) = 0,8 \quad \wedge \quad P(X \leq 12) = 0,9$. Estandarizamos y nos queda que:

$$\Phi\left(\frac{11 - \mu}{\sigma}\right) = 0,8 \quad \wedge \quad \Phi\left(\frac{12 - \mu}{\sigma}\right) = 0,9$$

De la tabla obtenemos que $F(0,8416) = 0,8 \quad \wedge \quad F(1,2816) = 0,9$. Planteamos:

$$\begin{cases} \frac{11 - \mu}{\sigma} = 0,8416 \\ \frac{12 - \mu}{\sigma} = 1,2816 \end{cases}$$

Resolvemos y obtenemos que $\mu = 9,09$ y $\sigma = 2,27$. Es decir: $X:N(9,09 ; 2,27)$.

Funciones lineales de variables aleatorias normales

Si X es una variable aleatoria normal $X:N(\mu_x ; \sigma_x)$

e Y es una función lineal de X , es decir, $Y = aX+b$ con $a,b \in \mathfrak{R}$,

entonces Y también es una variable aleatoria normal $Y:N(\mu_y ; \sigma_y)$

y sus parámetros valen:

$$\mu_y = a \mu_x + b$$

$$\sigma_y = \sigma_x |a|$$

La demostración (queda para el lector) consiste en hacer el cambio de variables $Y = aX+b$ y encontrar la distribución de Y .

Ejemplo:

El plástico de una botella de $2\frac{1}{4}$ litros cuesta 30 centavos. La gaseosa cuesta 40 centavos por litro. La cantidad de gaseosa (en litros) que se envasa en la botella es $N(2; 0,1)$.

¿Cuál es la probabilidad de que el costo total de una botella sea menor a 1,20 pesos?

$$Y = 40 X + 30 \Rightarrow Y: N(110; 4)$$

$$P(Y < 120) = F(120) = \Phi(2,5) = 0,99379$$

La estandarización como función lineal:

La estandarización $Z = \frac{X - \mu_x}{\sigma_x}$ también podría ser escrita como $Z = \frac{1}{\sigma_x} X - \frac{\mu_x}{\sigma_x}$, con lo cual vemos que es un caso particular de función lineal, en el cual $a = 1/\sigma_x$ y $b = -\mu_x/\sigma_x$. Entonces:

$$\mu_z = a \mu_x + b = (1/\sigma_x) \mu_x + -\mu_x/\sigma_x = 0$$

$$\sigma_z = \sigma_x \cdot |a| = \sigma_x \cdot 1/\sigma_x = 1$$

Con lo cual verificamos que del cambio de variables que usamos para estandarizar efectivamente resulta una normal $N(0;1)$.

Suma de variables aleatorias normales independientes

Si X e Y son normales e independientes, su suma también es normal:

Si:

$$X: N(\mu_x; \sigma_x)$$

$$Y: N(\mu_y; \sigma_y)$$

X, Y independientes

$$Z = X + Y$$

entonces:

$$Z : N \left(\mu_z = \mu_x + \mu_y \quad ; \quad \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2} \right)$$

Combinación lineal de variables aleatorias normales independientes

Si se tienen n variables aleatorias normales X_i , cada una con su propia media y varianza, y todas independientes entre sí, entonces la combinación lineal de esas variables también es una variable aleatoria normal:

Si:

$$X_i: N(\mu_i; \sigma_i) \quad \text{con } i = 1, 2, \dots, n$$

todas las X_i independientes

$$Z = \sum_{i=1}^n \alpha_i X_i$$

entonces:

$$Z : N \left(\mu_z = \sum_{i=1}^n \alpha_i \mu_i \quad ; \quad \sigma_z = \sqrt{\sum_{i=1}^n \alpha_i^2 \sigma_i^2} \right)$$

Ejemplo:

El plástico cuesta 0,5 centavos por gramo. La gaseosa cuesta 40 centavos por litro. La cantidad de plástico necesario para hacer una botella de $2^{1/4}$ litros es, en gramos, $N(100;10)$. La cantidad de gaseosa (en litros) que se envasa en la botella es $N(2;0,1)$. ¿Cuál es la probabilidad de que el costo total de una botella sea menor a 1,20 pesos?

X = gramos de plástico usados

Y = litros de coca cola embotellados

Z = costo total de una botella

$$Z = 0,5 X + 40 Y$$

Z: $N(130 ; 6,4)$

$$P(Z < 120) = \Phi(-1,56) = 0,059$$

Cuidado (Errores habituales)

1) **Sumar n variables aleatorias no es lo mismo que multiplicar por n una variable aleatoria.**

Por ejemplo, no es lo mismo tomar el peso de una docena de huevos, que tomar el peso de un huevo y multiplicarlo por 12.

Al calcular el peso de una docena de huevos, se están sumando 12 variables aleatorias independientes. Al multiplicar el peso de un huevo por 12, se está multiplicando por 12 una sola variable aleatoria. Es decir:

$$Y = \text{peso de 12 huevos} = X_1 + X_2 + \dots + X_{12}$$

$$Z = 12 \text{ veces el peso de un huevo} = 12 X$$

Veamos cómo quedan distribuidas:

Y es una combinación lineal de 12 variables aleatorias independientes. Los α de la combinación lineal valen todos 1. Los 12 huevos están distribuidos idénticamente, con lo cual $\mu_{x_i} = \mu_x$, $\sigma_{x_i} = \sigma_x$.

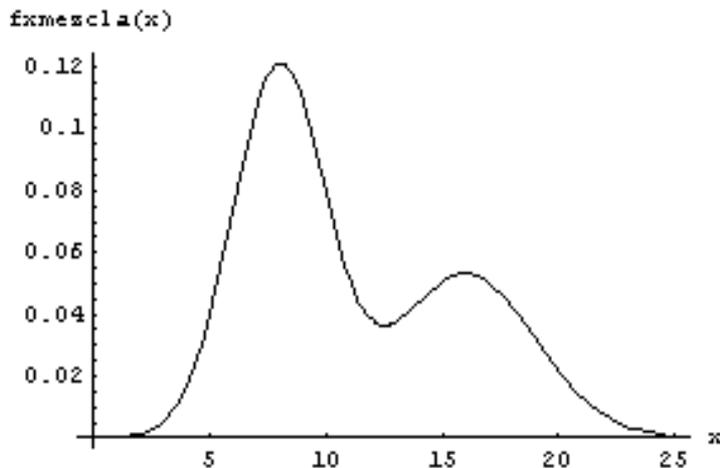
$$Y : N \left(\mu_y = \sum_{i=1}^{12} \alpha_i \mu_i = 12 \mu_x \quad ; \quad \sigma_y = \sqrt{\sum_{i=1}^{12} \alpha_i^2 \sigma_i^2} = \sqrt{12} \sigma_x \right)$$

Z es una función lineal de X, es decir $Z = 12 X$. También la podemos ver como una combinación lineal de un solo término. Usando la fórmula de la lineal, queda $a=12$, $b=0$.

$$Z: N(\mu_z = a \mu_x + b = 12 \mu_x ; \sigma_z = \sigma_x |a| = 12 \sigma_x)$$

Vemos que las medias nos quedaron iguales, pero el desvío de la suma de 12 huevos nos quedó menor. ¿Por qué sucede esto? El desvío (y la varianza) son una medida de cuánto tienden a alejarse de la media los valores de la variable aleatoria. Como los valores pueden estar a la izquierda o a la derecha de la media, sumando n variables aleatorias las distancias a la media de cada uno de esos n valores tienden a compensarse, por eso la varianza de la suma de n variables es menor que la de la multiplicación de una variable por n .

2) La mezcla de variables aleatorias normales NO resulta una variable normal.



Si bien la combinación lineal de variables aleatorias normales es normal, la mezcla no es una combinación lineal de variables aleatorias sino de las funciones de densidad de dichas variables aleatorias. Y la combinación lineal de 2 o más de funciones de densidad de variables normales no resulta una función de densidad de variable normal.

En el gráfico se ve un ejemplo de la mezcla de 2 variables aleatorias normales:

$$X:N(16;3)$$

$$Y:N(8;2)$$

$$P(X) = \frac{2}{5}$$

$$P(Y) = \frac{3}{5}$$

Como vemos, la variable aleatoria mezcla de esas dos normales dista mucho de ser normal.

Problemas típicos

1) El consumo de una determinada máquina por día, medido en kwh, es una V.A. normal con media 30 y varianza 100. Calcule la probabilidad de que en un determinado día, la máquina consuma:

- a) menos de 50 kwh
- b) menos de 30 kwh
- c) menos de 23 kwh

d) entre 30 y 40 kwh

Resolución:

X = consumo de la máquina en un día determinado

$\Rightarrow X:N(30;10)$ (no olvidar que el segundo parámetro de la distribución normal es el desvío, es decir, no la varianza sino su raíz)

$$P(X < 50) = F_X(50) = \Phi\left(\frac{50 - 30}{10}\right) = \Phi(2) = 0,97725$$

a)

b) La campana de Gauss es simétrica respecto de la media, por lo cual la probabilidad de que la variable sea menor que la media es la mitad del área total, es decir, la mitad de 1. Entonces $P(X < 30) = 0,5$

c)

$$P(X < 23) = F_X(23) = \Phi\left(\frac{23 - 30}{10}\right) = \Phi(-0,7) = 1 - \Phi(0,7) = 1 - 0,75804 = 0,24196$$

d)

$$P(30 < X < 40) = F_X(40) - F_X(30) = \Phi\left(\frac{40 - 30}{10}\right) - \Phi\left(\frac{30 - 30}{10}\right) = \Phi(1) - \Phi(0) = 0,84134 - 0,5 = 0,34134$$

2) El consumo de una determinada máquina por día, medido en kwh, es una V.A. normal con media 30 y varianza 100. Nos preguntan cuál es el máximo consumo que la máquina puede tener. Indique cuál debe ser la respuesta, si queremos que nuestra predicción del máximo se cumpla:

a) el 95% de los días.

b) el 50% de los días.

c) el 20% de los días.

Resolución:

X = consumo de la máquina en un día determinado

$\Rightarrow X:N(30;10)$

Debemos encontrar x tal que $P(X \leq x)$ sea la probabilidad dada

$$P(X \leq x) = 0,95 \Rightarrow \Phi\left(\frac{x - 30}{10}\right) = 0,95$$

a)

de la tabla conseguimos que el fractil $z_{0,95} = 1,645$. Es decir:

$$\Phi(1,645) = 0,95 \Rightarrow \frac{x - 30}{10} = 1,645 \Rightarrow x = 46,45$$

b) Como la distribución es simétrica, si $P(X \leq x) = 0,5$ entonces $x = \mu = 30$

$$P(X \leq x) = 0,2 \Rightarrow \Phi\left(\frac{x - 30}{10}\right) = 0,2$$

c)

Si tenemos en nuestra tabla el fractil $z_{0,2}$ procedemos como en a. Pero si nuestra tabla solo tiene la mitad de la distribución, debemos recordar que, por simetría:

$$\Phi(-z) = 1 - \Phi(z)$$

con lo cual

$$\Phi\left(\frac{x-30}{10}\right) = 0,2 \Rightarrow 1 - \Phi\left(\frac{30-x}{10}\right) = 0,2 \Rightarrow \Phi\left(\frac{30-x}{10}\right) = 0,8$$

y luego buscamos en la tabla de fractiles $z_{0,8}$ y procedemos como en a:

$$\Phi(0,842) = 0,8 \Rightarrow \frac{30-x}{10} = 0,842 \Rightarrow x = 21,58$$

3) El consumo de una determinada máquina por día, medido en kwh, es una V.A. normal. El 30% de los días consume menos de 10 kwh, y el 80% de los días consume menos de 60 kwh. ¿Cuál es la media y la varianza de la distribución?

Resolución:

Si X es el consumo en kwh por día de la máquina, entonces los datos que nos están dando son:

$$P(X \leq 10) = 0,3$$

$$P(X \leq 60) = 0,8$$

Es decir:

$$F_x(10) = 0,3$$

$$F_x(60) = 0,8$$

Lo cual, como X es normal, equivale a:

$$\Phi\left(\frac{10-\mu}{\sigma}\right) = 0,3$$

$$\Phi\left(\frac{60-\mu}{\sigma}\right) = 0,8$$

Donde μ y σ son los parámetros que desconocemos. Ahora buscamos en la tabla los fractiles $z_{0,3}$ y $z_{0,8}$. Al igual que sucedía con Φ para valores negativos, puede ser que no tengamos en la tabla el valor de los fractiles de menos de 0,5 por lo cual podemos no tener el fractil $z_{0,3}$. En ese caso recordemos que $z_{0,3}$ es en realidad z tal que $\Phi(z) = 0,3$.

$\Phi(z) = 1 - \Phi(-z)$, de donde vemos que $z_{0,3}$ también es z tal que $\Phi(-z) = 0,7$.

Entonces $-z = 0,5244$, es decir, $z_{0,3} = -0,5244$. $z_{0,8}$ siempre figura en la tabla, y vale 0,8416. Luego:

$$\frac{10-\mu}{\sigma} = -0,5244$$

$$\frac{60-\mu}{\sigma} = 0,8416$$

Nos quedan dos ecuaciones con dos incógnitas, de donde podemos despejar que:

$$\mu = 25,55$$

$$\sigma = 29,65$$

4) El chocolate tiene una densidad de 3g /cm³. El molde que se utiliza para fabricar barras de chocolate produce barras cuyo volumen en cm³ está distribuido normalmente con media 30 y desvío 5. Si la caja pesa 25g, ¿cuál es la probabilidad de que una caja de chocolate pese menos de 120 gramos?

Resolución:

Si X es el peso de chocolate, nos dicen que $X:N(30;5)$

Si Y es el peso de la caja de chocolate, tenemos que $Y = 3X + 25$.

Vemos que Y es una función lineal de una variable aleatoria normal.

Nos piden $P(Y < 120)$

Hay 2 formas de resolver este problema. Podemos usar el teorema que nos da la distribución de una función de una variable aleatoria normal, según el cual $Y=aX+b$ resulta ser una variable normal, con:

$$\mu_Y = a \mu_X + b = 115$$

$$\sigma_Y = |a|\sigma_X = 15$$

De donde luego:

$$P(Y < 120) = F_Y(120) = \Phi\left(\frac{120 - 115}{15}\right) = \Phi(0,33) = 0,63$$

Pero también podríamos haber trabajado directamente con la expresión de Y en términos de X, es decir:

$$P(Y < 120) = P(3X + 25 < 120) = P(X < 31,67) = F_X(31,67) = \Phi\left(\frac{31,67 - 30}{5}\right) = \Phi(0,33) = 0,63$$

Y de esa forma obtenemos el mismo resultado.

5) Carl Lewis puede correr los 100 metros llanos en un tiempo distribuido normalmente $N(7;3)$ en segundos. Su rival Ben Johnson puede hacer esa misma distancia en un tiempo distribuido normalmente según $N(9;2)$ en segundos.

- ¿Cuál es la probabilidad de que Carl Lewis le gane a Ben Johnson?**
- ¿Cuál es la probabilidad de que le gane aunque le de 1 segundo de ventaja?**

Resolución:

Si X e Y son los tiempos que tardan Carl Lewis y Ben Johnson respectivamente, entonces:

$$X:N(7;3)$$

$$Y:N(9;2)$$

$$a) P(\text{gane Carl Lewis}) = P(X < Y) = P(X - Y < 0)$$

Si tomamos $Z = X - Y$, y consideramos que los tiempos que tardan los dos atletas

son independientes, entonces podemos usar:

$$Z : N \left(\mu_z = \sum_{i=1}^n \alpha_i \mu_i \quad ; \quad \sigma_z = \sqrt{\sum_{i=1}^n \alpha_i^2 \sigma_i^2} \right)$$

Con lo cual queda:

$$Z: N(-2; 3,6)$$

Luego:

$$P(Z < 0) = F_Z(0) = \Phi \left(\frac{0 - (-2)}{3,6} \right) = \Phi(0,56) = 0,71$$

b) P(gane Carl Lewis dando 1 segundo de ventaja) = P(X+1 < Y) = P(X - Y < -1)

$$P(Z < -1) = F_Z(-1) = \Phi \left(\frac{-1 - (-2)}{3,6} \right) = \Phi(0,28) = 0,61$$

6) Cada 100g, el dulce de leche tiene 300 calorías y el flan 180 calorías. En un flan con dulce de leche la cantidad de flan en gramos es N(50;10) y la cantidad de dulce de leche en gramos es N(25;15).

a) ¿Cuál es la probabilidad de que un flan con dulce de leche tenga menos de 220 calorías?

b) Si una persona, en vez de flan con dulce de leche, come acelga(10cal/100g), en una cantidad distribuida normalmente con media 2kg y desvío 50g, ¿cuál es la probabilidad de que la que come acelga ingiera más calorías que una que come flan con dulce de leche?

Resolución:

a) Si llamamos X a la cantidad de flan e Y a la cantidad de dulce de leche, entonces las calorías de un flan con dulce de leche son:

$$C = 1,8 X + 3 Y$$

donde:

$$X: N(50;10)$$

$$Y: N(25;15)$$

Luego como C es una combinación lineal de variables normales independientes, entonces C también es una variable normal, y vale:

$$C : N \left(\mu_c = \sum_{i=1}^n \alpha_i \mu_i \quad ; \quad \sigma_c = \sqrt{\sum_{i=1}^n \alpha_i^2 \sigma_i^2} \right)$$

Luego:

$$\mu_c = 1,8 \mu_x + 3 \mu_y = 1,8 \cdot 50 + 3 \cdot 25 = 165$$

$$\sigma_c = \sqrt{1,8^2 \cdot \sigma_x^2 + 3^2 \cdot \sigma_y^2} = \sqrt{1,8^2 \cdot 10^2 + 3^2 \cdot 15^2} = 48,5$$

Y nos piden:

$$P(C < 220) = F_C(220) = \Phi\left(\frac{220 - 165}{48,5}\right) = \Phi(1,13) = 0,87$$

b) Si Z es la cantidad de acelga ingerida por la otra persona, entonces $Z: N(2000; 60)$ y la cantidad de calorías que ingiere es $A = 0,1 Z$. Nos piden $P(A > C)$, es decir, $P(A - C > 0)$. Tenemos varios caminos para resolverlo. Por ejemplo, podemos encontrar la distribución de A, y luego encontrar la distribución de la resta de A y C. También podemos dejar A en función de Z y C en función de X e Y, y encontrar la distribución de la siguiente combinación lineal:

$$W = 0,1A - 1,8 X - 3Y$$

Como la distribución de C ya la tenemos, el camino que nos conviene a los efectos de hacer la menor cantidad de cuentas posible es encontrar la distribución de:

$$W = 0,1.Z - C$$

W es una variable aleatoria normal con:

$$\mu_w = 0,1 \mu_z + (-1) \mu_c = 0,1 \cdot 2000 - 165 = 35$$

$$\sigma_w = \sqrt{0,1^2 \cdot \sigma_z^2 + (-1)^2 \cdot \sigma_c^2} = \sqrt{0,1^2 \cdot 60^2 + (-1)^2 \cdot 48,5^2} = 48,8$$

Luego la probabilidad de que la persona que come acelga ingiera más calorías que la que come flan con dulce de leche vale:

$$P(A > C) = P(W > 0) = 1 - P(W < 0) = 1 - F_w(0) = 1 - \Phi\left(\frac{0 - 35}{48,8}\right) = 1 - \Phi(-0,72) = \Phi(0,72) = 0,76$$

7) El peso de una naranja está distribuido normalmente según $N(0,1 ; 0,015)$ en kg.

a) ¿Cuál es la probabilidad de que 100 naranjas pesen menos de 12 kg.?

b) Si las naranjas aportan 100 calorías por kg., ¿cuál es la probabilidad de que una naranja aporte menos de 12 calorías?

Resolución:

Este ejemplo muestra la diferencia entre sumar n variables aleatorias idénticamente distribuidas y multiplicar una variable aleatoria por n.

Tomando X como el peso de una naranja queda $X: N(0,1 ; 15)$

Llamaremos Y al peso de 100 naranjas, y Z a la cantidad de calorías aportadas por una naranja.

$$Y = X_1 + X_2 + \dots + X_n = \sum_{i=1}^{100} X_i$$

a)

Con lo cual Y es una combinación lineal de 100 variables aleatorias independientes.

Los α de la combinación lineal valen todos 1. Los 100 huevos están distribuidos idénticamente, con lo cual $\mu_{x_i} = \mu_x$, $\sigma_{x_i} = \sigma_x$. Además asumiremos que los pesos de las naranjas son independientes.

$$Y : N\left(\mu_y = \sum_{i=1}^{100} \alpha_i \mu_i = 100 \mu_x = 10 \quad ; \quad \sigma_y = \sqrt{\sum_{i=1}^{100} \alpha_i^2 \sigma_i^2} = \sqrt{100 \cdot \sigma_x^2} = 1,22\right)$$

Luego:

$$P(Y < 12) = F_Y(12) = \Phi\left(\frac{12-10}{1,22}\right) = \Phi(1,63) = 0,95$$

b) $Z = 100X$.

Podemos verla como una función lineal de X (con término independiente $b=0$) o como una combinación lineal de una sola variable. De cualquiera de las dos maneras, resulta:

$$Z : N(\mu_z = 100\mu_x = 10 \quad ; \quad \sigma_z = 100\sigma_x = 1,5)$$

Luego:

$$P(Z < 12) = F_Z(12) = \Phi\left(\frac{12-10}{1,5}\right) = \Phi(1,33) = 0,91$$

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004

Teorema central del límite

Si \bar{X} es el promedio de una muestra de tamaño n de una población con media μ y

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

desvío estándar σ , entonces la variable aleatoria Z tiene una distribución aproximadamente normal estándar, bajo las siguientes condiciones:

- Si $n > 30$, la distribución de Z es aproximadamente normal estándar sin importar la distribución de las x .
- Si $n \leq 30$, la distribución de Z es aproximadamente normal solamente si la distribución de las x no difiere mucho de la distribución normal (por ejemplo: si es simétrica).
- Si la distribución de las x es normal, la distribución de Z es normal sin importar el valor de n .

Distribución de la suma de variables aleatorias

Si se tienen n variables aleatorias independientes e idénticamente distribuidas, y el valor de n cumple con las condiciones enunciadas más arriba, el teorema central del límite permite hallar la distribución de la suma de dichas variables, de la siguiente manera:

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

Si a la suma de las X_i (la variable cuya distribución queremos encontrar) la llamamos Y , entonces queda:

$$\bar{X} = \frac{Y}{n}$$

Reemplazándolo en la Z dada por el teorema central del límite, queda:

$$Z = \frac{\frac{Y}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\frac{1}{n}(Y - n\mu)}{\frac{\sigma}{\sqrt{n}}} = \frac{Y - n\mu}{\sqrt{n}\sigma}$$

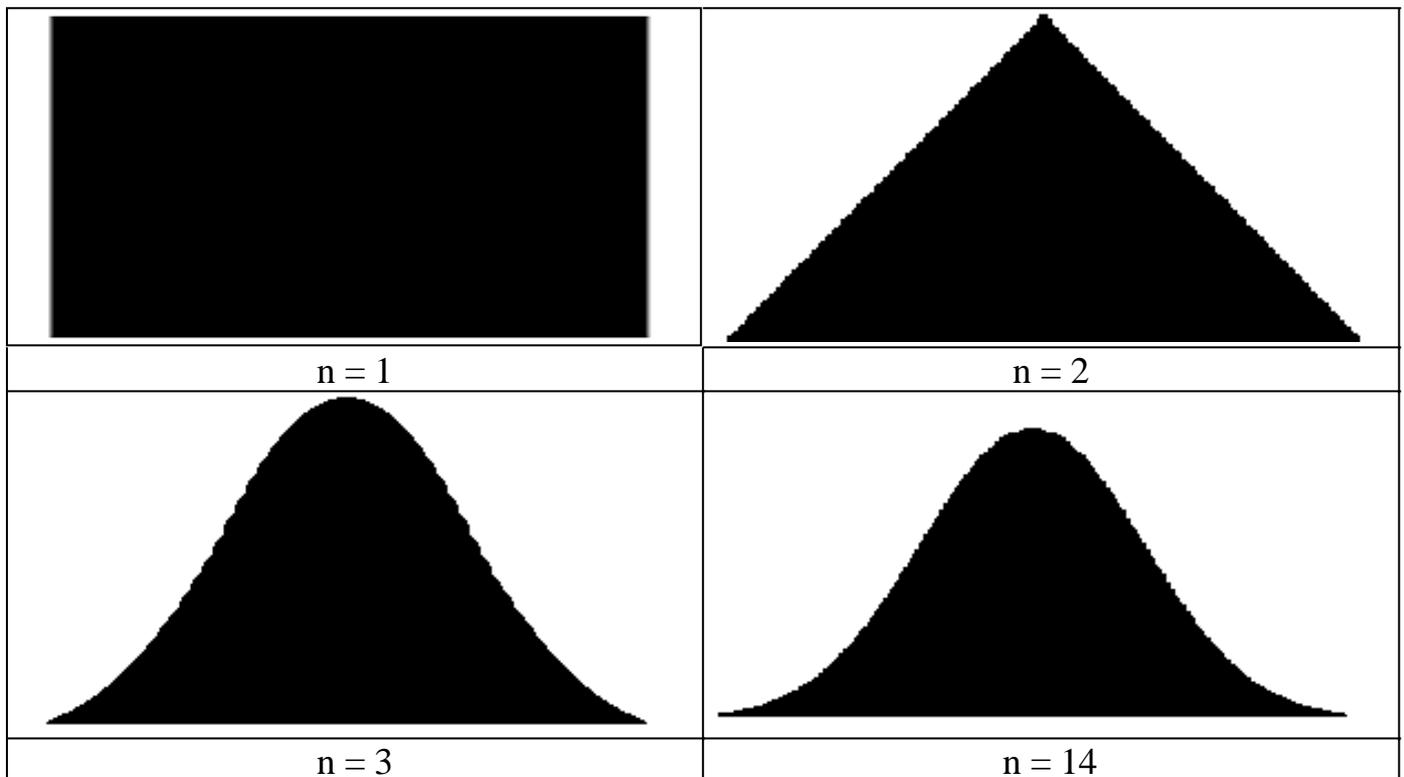
Lo que quedó lo podemos ver como la estandarización de una cierta variable normal Y (en realidad es aproximadamente normal). Si "desestandarizamos" nos queda que Y , es decir, **la suma de todas las variables X_i , es una variable**

normal con: $\mu_Y = n\mu$; $\sigma_Y = \sqrt{n} \sigma$

Este es el mismo resultado que habíamos obtenido para la suma de normales, con la diferencia de que ahora tenemos la condición de que n debe ser lo suficientemente grande.

En conclusión, la suma de una determinada cantidad de variables aleatorias independientes e idénticamente distribuidas resulta ser una variable normal en caso de que las variables sean normales, y aproximadamente normal en caso de que no sean normales pero n sea lo suficientemente grande.

A continuación vemos la forma de la distribución de la suma de n variables uniformes, para varios valores de n :



Suma de variables uniformes

En la práctica, la suma de 4 variables aleatorias uniformes independientes e idénticamente distribuidas se considera aproximadamente normal.

Problemas típicos

1) El peso en kg. de cada pieza es una variable aleatoria X distribuida según:

$$f_X(x) = \begin{cases} \frac{1}{4} & 1 < x < 5 \\ 0 & \forall \text{ otro } x \end{cases}$$

¿Cuál es la media y la varianza del peso de un lote de 100 piezas?

Resolución:

El peso de cada pieza es una variable aleatoria X_i . Todas esas X_i están idénticamente distribuidas según la f dada y se suponen independientes. Entonces la suma del peso de 100 piezas, como $100 \geq 30$, es aproximadamente una variable aleatoria normal con media $100\mu_x$ y desvío $10\sigma_x$.

A partir de la distribución que nos dan para las X_i , calculamos que:

$$\mu_x = 3 \quad ; \quad \sigma_x = 4/3$$

Si llamamos Y al peso del lote de 100 piezas, entonces la media de Y es $\mu_Y = 100\mu_x = 300$, y la varianza de Y es $\sigma_Y^2 = (10\sigma_x)^2 = 178$.

2) El tiempo que se tarda en llevar a cabo una operación es una variable aleatoria con media = 10 minutos y desvío = 2 minutos.

a) ¿Cuál es la probabilidad de que se tarde menos de 9 horas en realizar 49 operaciones?

b) ¿Cuál es la probabilidad de que el tiempo promedio por operación sea menor a 9 minutos?

Resolución:

a) El tiempo que se tarda en llevar a cabo cada operación es una variable aleatoria X_i con media 10 minutos y desvío 4 minutos. La suma de los tiempos de 49 de esas operaciones es una variable aleatoria aproximadamente normal con media $49\mu_x = 490$ minutos y desvío $7\sigma_x = 28$ minutos. Si a dicha suma la llamamos Y , nos están pidiendo:

$$P(Y < 9 \text{ horas}) = P(Y < 540 \text{ minutos}) = F_Y(540) = \Phi\left(\frac{540 - 490}{28}\right) = \Phi(1.79) = 0.963$$

b) Podemos hacerlo de dos formas:

- La primera es usando la versión enunciada del teorema central del límite, que nos dice que si tenemos $n \geq 30$ variables aleatorias independientes e idénticamente distribuidas, entonces la distribución de su promedio es aproximadamente normal

con media μ y desvío $\frac{\sigma}{\sqrt{n}}$ donde μ y σ son la media y el desvío de las variables que estamos promediando. Entonces:

$$P(X < 9 \text{ minutos}) = F_X(9) = \Phi\left(\frac{9 - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = \Phi(-1.75) = 1 - \Phi(1.75) = 0.04$$

- La otra forma consiste en ver que si Y (la suma) sigue una distribución normal,

entonces Y/n (el promedio) también sigue una distribución normal, porque es una constante $(1/n)$ multiplicada por una variable normal. Según estudiamos en este capítulo, $W=Y/n$ tiene una distribución normal con media μ_Y/n y desvío σ_Y/n . Luego la probabilidad de que W sea menor a 9 minutos es:

$$P(W < 9 \text{ minutos}) = F_Z(9) = \Phi\left(\frac{9 - \frac{\mu_Y}{n}}{\frac{\sigma_Y}{n}}\right) = \Phi(-1.75) = 1 - \Phi(1.75) = 0.04$$

3) La NASA está planeando una misión tripulada a la Luna. La duración de cada tanque de oxígeno es una variable aleatoria con media 6 horas y desvío 1 hora. ¿Cuántos tanques se deben llevar, para que la probabilidad de que alcance el oxígeno para una misión de 10 días sea del 99,9%?

Resolución:

Para una misión de 10 días se necesitan 240 horas de oxígeno. Tenemos que ver cuántos tanques hay que sumar para que la probabilidad de que superen 240 horas sea 0,999. Para sumar las duraciones de los tanques, vamos a usar el teorema central del límite. Pero es necesario destacar que una de las condiciones del teorema central del límite es que la cantidad de variables que se suman sea 30 ó más. Y en este caso desconocemos la cantidad de variables que estamos sumando (justamente es eso lo que queremos averiguar). Observemos que $240 / 6 = 40$, con lo cual es medianamente razonable suponer que la cantidad de tanques que sumaremos será mayor a 30.

$$Y = \sum_{i=1}^n X_i$$

Aclarado esto, si $Y = \sum_{i=1}^n X_i$ donde las X_i son las duraciones de los tanques, e Y es lo suficientemente grande (suponemos que se cumple) entonces Y es

aproximadamente una variable aleatoria normal con media $6n$ y desvío $1\sqrt{n}$.

Planteamos:

$$P(Y > 240) = 0,999$$

$$1 - \Phi\left(\frac{240 - 6n}{\sqrt{n}}\right) = 0,999$$

$$\Phi\left(\frac{240 - 6n}{\sqrt{n}}\right) = 0,001$$

$$\frac{240 - 6n}{\sqrt{n}} = z_{0,001} = -3,09$$

Resolviendo obtenemos que n debería valer 43,39. Si n pudiera valer ese número, la probabilidad sería exactamente 0,999. Pero n debe ser un número entero. Si redondeamos para abajo, la probabilidad de que el oxígeno alcance sería menor a 0,999, y eso no cumple con lo pedido. Por lo tanto, debemos redondear para arriba y responder que hay que llevar 44 tanques.

Ese resultado verifica $44 \geq 30$ con lo cual fue legítimo aplicar el teorema central del

límite.

4) Un programador se alimenta a base de chocolates. Su consumo diario de calorías es una variable aleatoria con media 2000 calorías y desvío 40 calorías. Si cada chocolate provee una cantidad de calorías que es una variable aleatoria con media 1000 calorías y desvío 25 calorías, ¿Cuántos chocolates debería comprarse para poder estar encerrado un mes programando un sistema y que la probabilidad de tener que salir a comprar más chocolates sea solamente 0.01?

Resolución:

Llamaremos X_i al consumo de calorías del i -ésimo día, y Y_j a aporte de calorías del j -ésimo chocolate.

$$A = \sum_{i=1}^{30} X_i$$

Por el teorema central del límite, si $A = \sum_{i=1}^{30} X_i$ entonces A es aproximadamente normal con media $30 \cdot 2000 = 60000$ y desvío $\sqrt{30} \cdot 40 = 219,09$.

Para hacer lo mismo con la cantidad de chocolates, debemos suponer que serán 30 o más chocolates. Como $60000/1000 = 60 \gg 30$, es bastante razonable suponer

$$B = \sum_{j=1}^n Y_j$$

que serán más de 30 chocolates, con lo cual si $B = \sum_{j=1}^n Y_j$ entonces B es aproximadamente normal con media $n \cdot 1000$ y desvío $\sqrt{n} \cdot 25$.

Luego podemos plantear que la probabilidad de que no le alcancen los chocolates debe ser 0.01, es decir:

$$P(A > B) = 0.01$$

$$P(A - B > 0) = 0.01$$

Si tomamos $C = A - B$, resulta que por ser combinación lineal de normales C también es normal. Calculamos sus parámetros:

$$\mu_C = \mu_A - \mu_B = 60000 - 1000n$$

$$\sigma_C = \sqrt{\sigma_A^2 + \sigma_B^2} = \sqrt{48000 + 625n}$$

Continuamos:

$$P(C > 0) = 0.01$$

$$1 - P(C < 0) = 0.01$$

$$1 - \Phi\left(\frac{0 - (60000 - 1000n)}{\sqrt{48000 + 625n}}\right) = 0.01$$

$$\frac{0 - (60000 - 1000n)}{\sqrt{48000 + 625n}} = z_{0.01} = -2,33$$

Resolviendo, obtenemos $n = 59.32$

Si n pudiera valer exactamente 59.32, entonces la probabilidad de que los chocolates no alcancen sería exactamente 0,01. A mayor n , menor probabilidad de que no alcancen, y a menor n , mayor probabilidad de que no alcancen. Si redondeáramos para abajo, la probabilidad sería mayor de 0.01, por lo tanto

debemos redondear para arriba y responder 60.

Vemos que además $60 \geq 30$, con lo cual fue lícito usar el teorema central del límite para sumar las calorías de los chocolates.

5) Una persona utiliza diariamente para lavar sus platos una cantidad de detergente que es una variable aleatoria con media 5 ml y desvío 1 ml. ¿Qué capacidad debería tener una botella de detergente para que la probabilidad de que le dure un mes sea del 98% ?

Resolución:

Si X es el consumo en un mes, debemos buscar C tal que se cumpla:

$$P(X < C) = 0,98$$

Si Y_i es el consumo de cada día, la suma del consumo de 30 días será

aproximadamente normal con media $30\mu_Y = 150$ y desvío $\sqrt{30} \sigma_Y = \sqrt{30}$. Luego:

$$P(X < C) = F_X(C) = \Phi\left(\frac{C - 150}{\sqrt{30}}\right) = 0,98$$

$$\frac{C - 150}{\sqrt{30}} = z_{0,98} = 2.0537$$

$$C = 161,2$$

Luego la botella debería tener al menos 161,2 ml.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004

Aproximación de Binomial y Poisson por Normal

Para calcular probabilidades de distribuciones discretas con números grandes, es preciso sumar muchos términos, lo cual puede resultar poco práctico. Sin embargo las características de algunas distribuciones, como la binomial y la Poisson, permiten muy buenas aproximaciones mediante la distribución normal. Y como la distribución normal se puede obtener de una tabla, el problema de sumar una gran cantidad de términos queda reducido a buscar uno o dos valores en una tabla. A continuación se presentan los métodos y justificaciones de cómo efectuar tales aproximaciones.

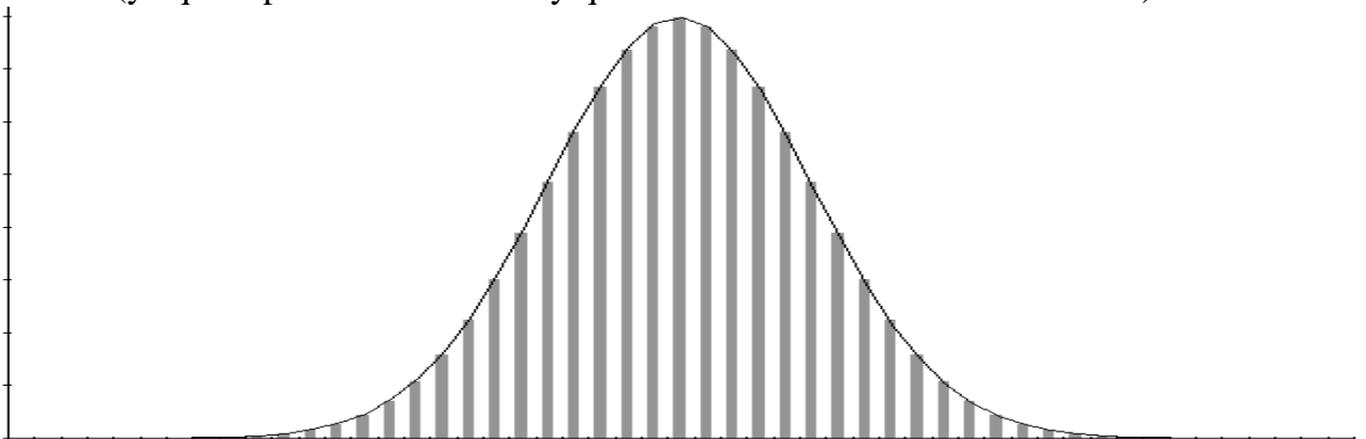
Aproximación de la distribución binomial por la distribución normal

Si X es una variable distribuida binomialmente, con $n \geq 10$ y p cercano a 0,5

$$Y = \frac{X - np}{\sqrt{np(1-p)}}$$

entonces la variable aleatoria Y tiene una distribución aproximadamente normal estándar.

Esto es válido porque si p es cercano a 0,5 y n es lo suficientemente grande (generalmente se pide $n \geq 10$) entonces la forma de la distribución binomial, a pesar de ser discreta, se parece mucho a la de la una distribución normal. El cambio de variable Y no es otra cosa que la estandarización de esa variable aproximadamente normal (ya que $n.p$ es la media de X y que el denominador es el desvío de X).



En el gráfico vemos una variable binomial($n = 100$; $p = 0,5$) junto con una variable normal($\mu = 50$; $\sigma = 5$).

Esta propiedad nos permite utilizar una variable normal estándar, que se encuentra tabulada, para ahorrarnos la engorrosa tarea de sumar una cantidad elevada de términos de probabilidades binomiales, especialmente cuando n es muy grande y la cantidad de éxitos está lejos de 0 y lejos de n , con lo cual la sumatoria tiene muchos términos aunque se intente restar del 1 en vez de sumar.

Queda por hacer una observación antes de poder utilizar esta propiedad. Al estar aproximando una distribución discreta por una continua, lo que se hace es tomar intervalos de la continua, que representan los valores puntuales de la discreta. Por ejemplo, consideraremos que la discreta vale 43, si la continua tiene cualquier valor entre 42,5 y 43,5. Entonces la probabilidad de que la discreta esté entre 8 y 12 no es la probabilidad de que la continua esté entre 8 y 12 sino de que esté entre 7,5 y 12,5. Considerar esto se conoce como "corrección por continuidad".

Ejemplo:

Se tiene una variable aleatoria $X:Bi(n = 50 ; p = 0,4)$. ¿Cuál es la probabilidad de que X sea menor a 20?

$$P(X < 20) = \sum_{x=0}^{19} P(X = x) = \sum_{x=0}^{19} \binom{n}{x} p^x (1-p)^{n-x}$$

Podríamos hacer . Esto
 demandaría sumar 20 términos, y arroja el resultado 0,44648

Sin embargo, y a menos que se necesite el resultado exacto, podemos usar la aproximación normal para resolver el problema. Estamos buscando $P(X < 20)$, lo cual es igual a:

$$P(0 \leq X \leq 19)$$

Hacemos la corrección por continuidad:

$$P(0 \leq X \leq 19) \cong P(-0,5 \leq X \leq 19,5)$$

Tomamos el cambio de variables:

$$Y = \frac{X - np}{\sqrt{np(1-p)}}$$

con lo cual Y tendrá una distribución aproximadamente normal estándar.

Dejamos X en función de Y :

$$X = \sqrt{np(1-p)} Y + np$$

Luego reemplazamos X por su definición en términos de Y en la probabilidad que estábamos buscando:

$$P(-0,5 \leq X \leq 19,5) = P\left(\frac{-0,5 - np}{\sqrt{np(1-p)}} \leq Y \leq \frac{19,5 - np}{\sqrt{np(1-p)}}\right) = P(-5,92 \leq Y \leq -0,14)$$

Lo cual, por propiedades de la función de distribución acumulada queda:

$$P(-5,92 \leq Y \leq -0,14) = F_Y(-0,14) - F_Y(-5,92)$$

Como estamos considerando a Y una normal estándar, entonces:

$F_Y(-0,14) - F_Y(-5,92) = \Phi(-0,14) - \Phi(-5,92) = (1 - \Phi(0,14)) - (1 - \Phi(5,92))$
 $= \Phi(5,92) - \Phi(0,14) = 1 - 0,55567 = 0,44433$
 Observemos que el resultado aproximado 0,44433 es prácticamente igual al resultado exacto 0,44648.

Demostración

Se provee esta demostración porque constituye un buen ejemplo de aplicación del teorema central del límite.

Si X es la cantidad de éxitos en una muestra en n experimentos de Bernoulli, entonces X es una variable aleatoria cuya distribución se conoce como binomial. Toda variable binomial es en esencia la suma de n variables de Bernoulli (unos y ceros). Como vimos para la distribución binomial:

$$E(X) = n.p$$

$$\sigma_x^2 = n.p.(1-p)$$

También vimos que, por el teorema central del límite, para n lo suficientemente grande, la suma de n variables tiene aproximadamente una distribución normal, con determinadas media y varianza. Particularmente cuando X es binomial, si

$np \geq 5$ y $n(1-p) \geq 5$ (lo cual también garantiza que p esté lo suficientemente alejada de 0 y 1 para que no se "aplaste") entonces su distribución se puede

aproximar por una normal, con media $n.p$ y desvío $\sqrt{n.p.(1-p)}$. Queda:
 $X: N(n.p; \sqrt{n.p.(1-p)})$ (aproximadamente).

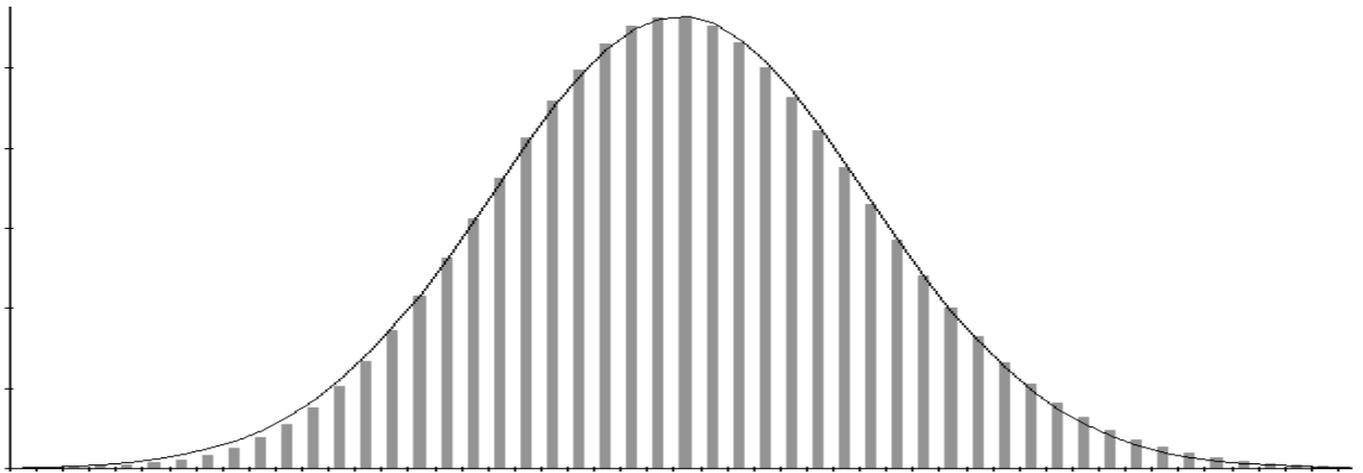
Luego, tomando el cambio de variables $Y = \frac{X - np}{\sqrt{np(1-p)}}$, Y tiene una distribución aproximadamente normal estándar.

Aproximación de la distribución de Poisson por la distribución normal

Si X es una variable de Poisson, con $\mu \gg 1$, entonces la variable aleatoria

$Y = \frac{X - \mu}{\sqrt{\mu}}$ tiene una distribución aproximadamente normal estándar.

Esto es válido porque si μ es mucho mayor que 1, entonces la forma de la distribución de Poisson, a pesar de ser discreta, se parece mucho a la de la una distribución normal. El cambio de variable Y no es otra cosa que la estandarización de esa variable aproximadamente normal (ya que μ es a la vez la media y la varianza de X)



En el gráfico vemos una variable de Poisson($\mu = 50$) junto con una variable normal($\mu = 50 ; \sigma = \sqrt{50}$).

Esta propiedad nos permite utilizar una variable normal estándar, que se encuentra tabulada, para ahorrarnos la engorrosa tarea de sumar una cantidad elevada de términos de probabilidades de Poisson al calcular probabilidades acumuladas, especialmente cuando necesitamos calcular la probabilidad acumulada para un valor que esté lejos del cero.

Queda por hacer una observación antes de poder utilizar esta propiedad. Al estar aproximando una distribución discreta por una continua, lo que se hace es tomar intervalos de la continua, que representan los valores puntuales de la discreta. Por ejemplo, consideraremos que la discreta vale 43, si la continua tiene cualquier valor entre 42,5 y 43,5. Entonces la probabilidad de que la discreta esté entre 8 y 12 no es la probabilidad de que la continua esté entre 8 y 12 sino de que esté entre 7,5 y 12,5. Considerar esto se conoce como "corrección por continuidad".

Ejemplo:

Se tiene una variable aleatoria $X: \text{Pois}(\mu = 60)$. ¿Cuál es la probabilidad de que X sea menor a 70?

$$P(X < 70) = \sum_{x=0}^{69} P(X = x) = \sum_{x=0}^{69} \frac{e^{-\mu} \mu^x}{x!}$$

Podríamos hacer . Esto demandaría sumar 70 términos, y arroja el resultado 0,88821.

Sin embargo, y a menos que se necesite el resultado exacto, podemos usar la aproximación normal para resolver el problema. Estamos buscando $P(X < 70)$, lo cual es igual a:

$$P(0 \leq X \leq 69)$$

Hacemos la corrección por continuidad:

$$P(0 \leq X \leq 69) \cong P(-0,5 \leq X \leq 69,5)$$

Tomamos el cambio de variables:

$$Y = \frac{X - \mu}{\sqrt{\mu}}$$

con lo cual Y tendrá una distribución aproximadamente normal estándar.

Dejamos X en función de Y :

$$X = \sqrt{\mu} Y + \mu$$

Luego reemplazamos X por su definición en términos de Y en la probabilidad que estábamos buscando:

$$P(-0,5 \leq X \leq 69,5) = P\left(\frac{-0,5 - \mu}{\sqrt{\mu}} \leq Y \leq \frac{69,5 - \mu}{\sqrt{\mu}}\right) = P(-7,81 \leq Y \leq 1,23)$$

Lo cual, por propiedades de la función de distribución acumulada queda:

$$P(-7,81 \leq Y \leq 1,23) = F_Y(1,23) - F_Y(-7,81)$$

Como estamos considerando a Y una normal estándar, entonces:

$$\begin{aligned} F_Y(1,23) - F_Y(-7,81) &= \Phi(1,23) - \Phi(-7,81) = \Phi(1,23) - (1 - \Phi(7,81)) \\ &= \Phi(1,23) + \Phi(7,81) - 1 = 0,89065 + 1 - 1 = 0,89065 \end{aligned}$$

Observemos que el resultado aproximado 0,89065 es prácticamente igual al resultado exacto 0,88821.

Problemas típicos

Deben considerarse modelos de problemas típicos los dos ejemplos dados en esta sección.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 6 de junio de 2004.

CAPÍTULO VII

Otras distribuciones particulares

Además de las distribuciones que hemos presentado en los capítulos sobre los procesos de Bernoulli y Poisson, y de la distribución normal, hay otras distribuciones interesantes y bastante comunes.

Distribución Multinomial

- Esta distribución es un caso general de la binomial, en el cual la cantidad de resultados posibles de cada experimento individual no es 2 (éxito y fracaso) sino k . Entonces tenemos que cada experimento arrojará 1 entre k resultados posibles E_1, E_2, \dots, E_k .
- p_1, p_2, \dots, p_k son las probabilidades de que salgan los resultados E_1, E_2, \dots, E_k respectivamente. Es decir, $p_i = P(E_i)$ con $i \in [1 ; k]$.
- El proceso consiste en hacer ese experimento n veces en forma independiente.
- Se toman las variables X_1, X_2, \dots, X_k como la cantidad de veces, dentro de las n , que sale el resultado E_1, E_2, \dots, E_k respectivamente.

$$P(X_1 = x_1 \cap X_2 = x_2 \cap \dots \cap X_k = x_k) = n! \prod_{i=1}^k \frac{p_i^{x_i}}{x_i!}$$

Observaciones

- Como siempre tiene que obligatoriamente salir uno de los resultados E_1, E_2, \dots, E_k , entonces

$$\sum_{i=1}^k p_i = 1 \quad \text{y además} \quad \sum_{i=1}^k X_i = n$$

Problemas típicos:

1) El 10% de los gatos que existen en una ciudad son rayados, el 30% son manchados, y el 60% son lisos. Si en un callejón de esa ciudad hay 10 gatos, cuál es la probabilidad de que haya:

a) 4 lisos, 3 manchados y 3 rayados.

b) 4 lisos y 3 manchados.

c) 4 lisos.

d) Sabiendo que hay 4 lisos, cuál es la probabilidad de que haya 3 manchados?

Resolución:

- Observamos que si el experimento consiste en observar un gato, hay $k = 3$ resultados posibles: que sea liso, que sea manchado y que sea rayado.
- Observamos que cada uno de los k resultados tiene una probabilidad asociada: $P(\text{liso}) = 0,6$; $P(\text{manchado}) = 0,3$; $P(\text{rayado}) = 0,1$. Dichas probabilidades suman 1.
- Observamos que el proceso va a consistir en observar $n = 10$ gatos independientes, y que la cantidad de gatos lisos más la cantidad de gatos manchados más la cantidad de gatos rayados va a sumar n .

Entonces las cantidades encontraremos de cada tipo de gato van a estar distribuidas multinomialmente, de la siguiente manera:

$$n = 10$$

Resultados posibles: liso, manchado, rayado.

$$P(\text{liso}) = p_L = 0,6$$

$$P(\text{manchado}) = p_M = 0,3$$

$$P(\text{rayado}) = p_R = 0,1$$

X_L : la cantidad de gatos lisos de los 10 que hay

X_M : la cantidad de gatos manchados de los 10 que hay

X_R : la cantidad de gatos rayados de los 10 que hay

Un ejemplo de este proceso podría ser:



En este ejemplo resultó ser $X_L = 6$, $X_M = 2$, $X_R = 2$.

Veamos ahora qué probabilidades nos piden:

a) "que haya 4 lisos, 3 manchados y 3 rayados"

$$P(X_L = 4 \cap X_M = 3 \cap X_R = 3) = \frac{10! \cdot 0,6^4 \cdot 0,3^3 \cdot 0,1^3}{4! \cdot 3! \cdot 3!} = 0,014697$$

b) "que haya 4 lisos y 3 manchados"

$$\sum_{i=1}^k X_i = n$$

No sabemos cuántos rayados, pero sabemos que es decir, que la suma de las cantidades parciales nos va a dar la cantidad total.

$$X_L + X_M + X_R = 10$$

$$X_R = 3$$

Además es fácil de ver, porque si en total hay 10, los siguientes sucesos:

- "que haya 4 lisos, 3 manchados y 3 rayados"
- "que haya 4 lisos y 3 manchados"

son equivalentes.

Por lo tanto la probabilidad es la misma que en el caso anterior: 0,014697

c) "que haya 4 lisos"

Acá ya no podemos aplicar lo mismo que en el ejemplo b, porque hay 2 cantidades indeterminadas.

Podríamos por ejemplo calcular esta probabilidad como la sumatoria de todas las probabilidades para $X_L = 4$ y todas las formas posibles de sumar 6 con X_M y X_R . Pero esto es obviamente poco práctico.

Una solución mejor sería inventar una nueva categoría de gato: "no liso". Y entonces tenemos una nueva multinomial con:

X_L : la cantidad de gatos lisos

X_N : la cantidad de gatos no lisos

$$p_L = 0,6$$

$$p_N = 0,4$$

Y luego calcular $P(X_L = 4 \cap X_N = 6)$ como vimos en la parte a.

Pero la mejor forma de ver el problema consiste en darse cuenta de que cada una de las X_L, X_M, X_R, X_N es en realidad una variable binomial, cuyo p es el p_i correspondiente y cuyo n es el n de la multinomial.

Entonces si tomamos por ejemplo el caso de X_L , vemos que puede ser considerado como una variable binomial con $n = 10$ y $p = 0,6$.

Y entonces la probabilidad de que en los 10 gatos haya 4 gatos lisos es:

$$p(X_L = 4) = \binom{n}{x_L} p_L^{x_L} (1 - p_L)^{n - x_L} = 0,11148$$

Además podemos extraer la conclusión de que la distribución binomial es, en realidad, una multinomial cuyo $k = 2$.

Si por ejemplo $X \sim \text{Bi}(n; p)$ entonces se podría definir una distribución multinomial equivalente con:

$$k = 2$$

$n = n$ (el mismo de la binomial)

Resultados posibles: E = éxito, F = fracaso

$$X_E = X$$

$$X_F = n - X$$

$$p_E = p$$

$$p_F = q = 1 - p$$

d) "Sabido que hay 4 lisos, cuál es la probabilidad de que haya 3 manchados?"

Aplicamos probabilidad condicional y obtenemos:

$$P\left(X_M = 3 \mid X_L = 4\right) = \frac{P(X_L = 4 \cap X_M = 3)}{P(X_L = 4)}$$

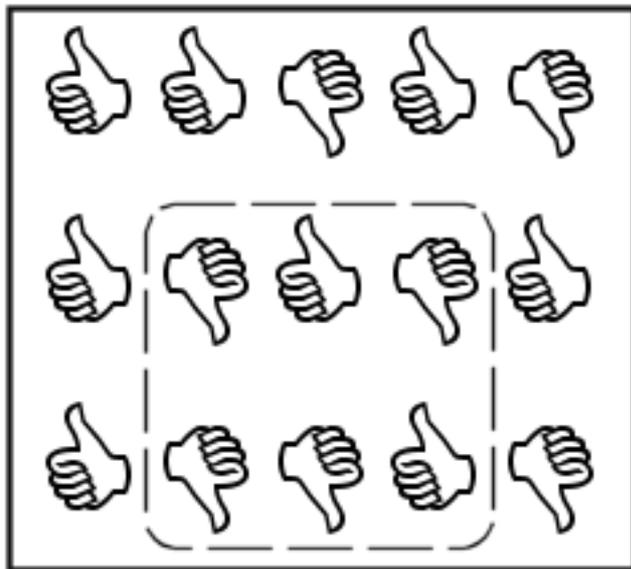
El numerador puede ser calculado con la distribución multinomial como fue visto en b.

El denominador puede ser calculado con la distribución binomial, como fue visto en c.

Hacemos la división y obtenemos:

$$P\left(X_M = 3 \mid X_L = 4\right) = 0,13184$$

Distribución Hipergeométrica



$$N = 15 \quad k = 8 \quad n = 6 \quad X = 2$$

Si se tienen N elementos,
de los cuales k son favorables,
y se extraen n elementos,
sin reposición,
y X es la cantidad de éxitos entre los n elementos que se se extraen,
entonces:

X:Hipergeom($N ; n ; k$)

$$P(X = x) = \frac{\binom{k}{x} \cdot \binom{N-k}{n-x}}{\binom{N}{n}}$$

con lo cual:

Problemas típicos:

- 1) Una caja contiene 12 bolitas, de las cuales 7 son negras, 3 son blancas y 2 son rojas. Si se sacan 4 bolitas al azar sin reposición, calcule la probabilidad de sacar:
 - a) alguna bolita roja

b) 4 bolitas negras

c) 4 bolitas negras, sabiendo que se sacaron bolitas negras.

Resolución:

a) vamos a considerar favorables a las bolitas rojas, y entonces tenemos que:

$N = 12$ hay 12 bolitas en total

$k = 2$ hay 2 bolitas favorables

$n = 4$ extraigo 4 bolitas

X: cantidad de bolitas rojas extraídas

$$P(\text{extraer alguna bolita roja}) = P(X > 0) = 1 - P(X \leq 0) = 1 - P(X = 0) = 0,57576$$

b) vamos a considerar favorables a las bolitas negras, y entonces tenemos que:

$N = 12$ hay 12 bolitas en total

$k = 7$ hay 7 bolitas favorables

$n = 4$ extraigo 4 bolitas

X: cantidad de bolitas negras extraídas

$$P(X = 4) = 0,0707$$

c) vamos a considerar favorables a las bolitas negras, y entonces tenemos que:

$N = 12$ hay 12 bolitas en total

$k = 7$ hay 7 bolitas favorables

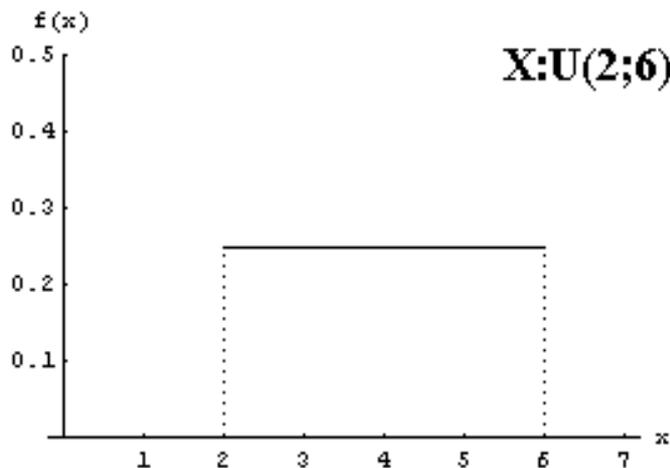
$n = 4$ extraigo 4 bolitas

X: cantidad de bolitas negras extraídas

$$P\left(\frac{X = 4}{X > 0}\right) = \frac{P(X = 4 \wedge X > 0)}{P(X > 0)} = \frac{P(X = 4)}{1 - P(X = 0)} = \frac{0,0707}{1 - 0,0101} = 0,07142857$$

Distribución Uniforme

Se dice que una variable aleatoria continua es uniforme entre a y b si el conjunto de sus valores posibles es el intervalo $[a;b]$ y todos esos valores tienen la misma probabilidad.



Variable aleatoria X uniforme entre 2 y 6

Dicho de otra forma, una variable aleatoria continua es uniforme entre a y b si su función de densidad es:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \forall \text{ otro } x \end{cases}$$

A primera vista podemos observar que tiene 2 parámetros: a y b . Estos 2 parámetros pueden tomar cualquier valor real ($a < b$) y al hacerlo definen una entre infinitas distribuciones uniformes posibles.

La notación $X:U(a;b)$ significa que la variable aleatoria X tiene una distribución uniforme con parámetros a y b , o dicho de otra forma, que la variable aleatoria X puede tomar cualquier valor al azar entre a y b , y todos esos valores posibles tienen la misma probabilidad de ocurrir.

Dicha característica de que todos los valores posibles tienen la misma probabilidad de ocurrir hace que esta distribución sea la ideal a suponer cuando sabemos que una variable aleatoria puede tomar cualquier valor dentro de un intervalo y no tenemos ninguna información acerca de las probabilidades relativas de ocurrir de los valores de ese intervalo. Por ejemplo si nos dicen que una naranja puede pesar entre 100 y 200 gramos, y no tenemos más información, supondremos que si X es el peso en gramos de una naranja cualquiera, entonces X está distribuida uniformemente entre 100 y 200, es decir $X:U(100;200)$.

Como para todas las distribuciones continuas, para calcular probabilidades podemos plantear:

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

que en este caso resulta valer:

$$P(X \leq x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & b < x \end{cases}$$

Aplicando análogamente las definiciones de media y varianza se obtiene:

$$E_x = \frac{a+b}{2} \quad \sigma_x^2 = \frac{(b-a)^2}{12}$$

Distribución Random

Cuando una variable aleatoria continua es uniforme entre 0 y 1, es decir, el caso particular de variable aleatoria uniforme con $a = 0$ y $b = 1$, entonces la distribución de la variable se conoce con el nombre de Random.

$X:\text{Random} \equiv X:U(0;1)$

La importancia de dicha distribución radica en sus aplicaciones en la simulación (apéndice C)

Problemas típicos

1) El peso de las naranjas está distribuido uniformemente entre 100 y 200 gramos.

a) ¿Cuál es la probabilidad de que una naranja pese menos de 180 gramos?

b) ¿Cuál es el valor esperado del peso de una naranja?

c) ¿Cuál es el valor más probable del peso de una naranja?

d) ¿Cuál es la varianza de la distribución?

Resolución:

$X:U(100;200)$

$$f_x(x) = \begin{cases} \frac{1}{100} & 100 \leq x \leq 200 \\ 0 & \forall \text{ otro } x \end{cases}$$

=>

$$P(X < 180) = \int_{-\infty}^{180} f_x(x) dx = \int_{100}^{180} \frac{1}{100} dx = \frac{1}{100} (180 - 100) = 0,8$$

a)

$$E_x = \frac{a + b}{2} = 150$$

b)

c) Vemos que todos los valores tienen la misma probabilidad, por lo tanto no hay un único "valor más probable". Entonces en este caso, el conjunto de valores más probables no es otra cosa que el conjunto de todos los valores posibles, es decir, el intervalo [100;200]

$$d) \sigma_x^2 = \frac{(b - a)^2}{12} = 833$$

2) Se sabe que la longitud de las varillas fabricadas por una máquina está distribuida uniformemente. Lo más corta que puede ser una varilla es 5 cm. Además se sabe que la probabilidad de que una varilla mida más de 15 cm es $\frac{1}{3}$.

a) Indique cuál es la función de densidad de la longitud de las varillas fabricadas por la máquina.

b) Si las varillas que miden menos de 8 cm se descartan, ¿cuál es la probabilidad de descartar una varilla?

Resolución:

a) Llamando X a la longitud de las varillas fabricadas por la máquina, entonces $X:U(a;b)$.

Nos dicen que el menor valor posible para X es 5. Entonces $a = 5$. Falta determinar b.

Nos dicen que:

$$P(X > 15) = \int_{15}^{+\infty} f_x(x) dx = 0,3333$$

Como $f_x(x)$ vale $\frac{1}{b-a}$ para $a < x < b$ y 0 para $x > b$, la integral queda:

$$\int_{15}^{+\infty} f_x(x) dx = \int_{15}^b \frac{1}{b-a} dx = \frac{1}{b-a} \int_{15}^b dx = \frac{b-15}{b-a} = \frac{b-15}{b-5} = 0,3333$$

Luego $b = 20$

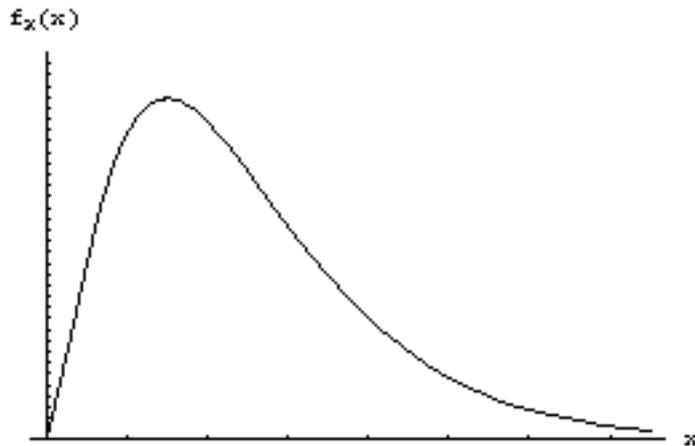
b) Nos están pidiendo $P(X < 8)$

$$\int_{15}^{+\infty} f_x(x) dx = \int_{15}^{+\infty} \frac{1}{b-a} dx = \frac{1}{b-a} \int_{15}^{+\infty} dx = \frac{b-15}{b-a} = \frac{b-15}{b-5} = 0,3333$$

Distribución Chi-cuadrada o Ji-cuadrada

o χ^2

La variable aleatoria X tiene la distribución chi-cuadrada o ji-cuadrada si su función de densidad de probabilidad es:



$$f_X(x) = \begin{cases} \frac{v^{\frac{v}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

donde

- el parámetro v es un número natural, y se conoce con el nombre de "grados de libertad"

- $\Gamma\left(\frac{v}{2}\right) = \int_0^{+\infty} x^{\frac{v}{2}-1} e^{-x} dx$

Notación

$X:\chi^2(v)$ se lee "X tiene una distribución ji-cuadrada con v grados de libertad"

Propiedades

Si

$$\mathbf{X}:\chi^2(v)$$

entonces:

- $E(X) = v$
- $\sigma_X^2 = 2v$

Relación con otras distribuciones

• Gamma:

Según se estudió en el capítulo anterior, la distribución gamma tiene los parámetros λ y k , donde λ es un número real positivo y k es un número natural.

Si hacemos $\lambda = 1/2$ y $k = n/2$, donde n es un número natural, obtenemos la distribución chi-cuadrada. Luego, la distribución chi-cuadrada es de la familia de la gamma, y por lo tanto tiene su misma forma, propiedades, etc.

Notemos que la distribución chi-cuadrada no es un caso particular de gamma, porque en la distribución gamma el parámetro k es un número natural, y en su lugar en la chi-cuadrada aparece $n/2$ con n natural, es decir: $1/2, 1, 3/2, 2, 5/2, 3$, etc.

Tampoco es un caso general de gamma, porque estamos fijando el parámetro λ , que en la gamma estaba libre, en $1/2$.

• Normal:

Si $X:N(0;1)$, y buscamos la distribución de $Y = X^2$, encontraremos que $Y:\chi^2(1)$. Es decir, el cuadrado de una normal estándar sigue la distribución chi-cuadrada con $v = 1$.

Propiedades reproductivas

Si tenemos

- m variables X_i
- $X_i:\chi^2(v_i)$
- X_i independiente de X_j para $i \neq j$

$$Y = \sum_{i=1}^m X_i$$

•

entonces:

- $Y:\chi^2(v_y)$

$$v_y = \sum_{i=1}^m v_i$$

•

Es decir, la suma de m variables chi-cuadradas independientes, resulta ser una variable chi-cuadrada cuyo ν es la suma de los ν de las variables originales.

Combinando esto último con la propiedad de que la distribución chi-cuadrada es la distribución del cuadrado de una variable normal estándar, se puede generalizar la relación entre la distribución chi-cuadrado y la distribución normal estándar:

Dadas Z_1, Z_2, \dots, Z_n , variables aleatorias normales estándar independientes, vale:

$$\text{Si } Y = Z_1^2 + Z_2^2 + \dots + Z_n^2 = \sum_{i=1}^n Z_i^2 \quad \text{entonces } Y : \chi^2(n)$$

Es decir, la suma de n variables normales estándar independientes al cuadrado, tiene una distribución chi-cuadrada con n grados de libertad.

Más aún, recordemos que una variable aleatoria normal cualquiera (no necesariamente estándar), se puede estandarizar restándole su media y dividiéndola por su desvío. Luego, dadas X_1, X_2, \dots, X_n , variables aleatorias independientes con distribución $N(\mu_i; \sigma_i)$, vale:

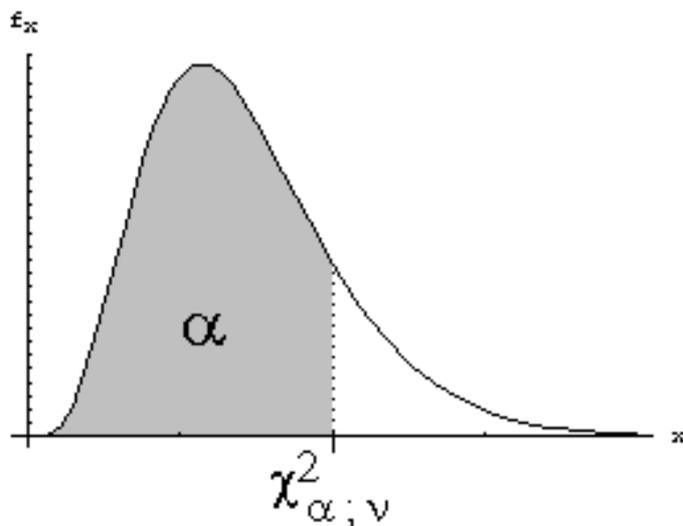
$$\text{Si } Y = \left(\frac{X_1 - \mu_1}{\sigma_1} \right)^2 + \dots + \left(\frac{X_n - \mu_n}{\sigma_n} \right)^2 = \sum_{i=1}^n \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \quad \text{entonces } Y : \chi^2(n)$$

Utilización

Debido al uso que le daremos (capítulo 8 en adelante), lo que nos interesa calcular de la distribución chi-cuadrada son sus fractiles. Es decir, los valores x tales que $P(X \leq x)$ es igual a un cierto α .

Sea $X: \chi^2(\nu)$, $\chi_{\alpha; \nu}^2$ es el valor x tal que $P(X \leq x) = \alpha$. Es decir, el valor tal que la probabilidad de que una variable chi-cuadrado con ν grados de libertad resulte menor que ese valor sea α . Dicho de otra forma, el valor que tiene un área α a la izquierda, bajo la curva de una chi-cuadrado con ν grados de libertad.

Dichos fractiles se encuentran tabulados en el apéndice D.



Sobre esto, una advertencia: muchos autores trabajan con el fractil de la chi-cuadrado a derecha en vez de a izquierda, y por eso muchas propiedades y tablas tienen las convenciones contrarias a las que utilizaremos. Recomendamos al consultar una tabla verificar previamente si los fractiles son a izquierda o a derecha.

Si debiéramos usar una tabla con los fractiles a derecha para buscar $\chi^2_{\alpha; \nu}$,

debemos buscar $\chi^2_{1-\alpha; \nu}$ ya que tener un área de α a la izquierda es lo mismo que tener un área de $1-\alpha$ a la derecha.

Aproximación

Para $\nu \geq 30$, se demuestra que si X es una variable chi-cuadrada con ν grados de libertad, entonces $Y = \sqrt{2X}$ es aproximadamente una normal $N(\sqrt{2\nu-1}; 1)$.

Es decir:

$$\left. \begin{array}{l} X : \chi^2(\nu) \\ \nu \geq 30 \\ Y = \sqrt{2X} \end{array} \right\} \Rightarrow Y : N(\sqrt{2\nu-1}; 1) \quad (\text{aprox.})$$

Consecuentemente, no vale realmente la pena tener tabulada la chi-cuadrada para $\nu \geq 30$, porque las probabilidades se pueden obtener con:

$$P(X \leq x) \approx \Phi(\sqrt{2x} - \sqrt{2\nu-1})$$

y los fractiles se pueden obtener con:

$$\chi^2_{\alpha; \nu} \approx \frac{(z_{\alpha} + \sqrt{2\nu-1})^2}{2}$$

donde Φ es la función de distribución acumulada a izquierda de la distribución normal estándar, y z_{α} es el fractil de la distribución normal estándar con un área de α a la izquierda. Tanto Φ como z_{α} se encuentran tabuladas.

Ejemplos

- Sea $X: \chi^2(13)$

- Se lee "X es una variable chi-cuadrada con 13 grados de libertad".

- El valor $\chi^2_{0.1;13}$:

- Es el fractil de la chi cuadrada para $\alpha = 0.1$ con 13 grados de libertad.
- Es decir: es el valor tal que hay probabilidad 0.1 de que una variable chi-cuadrada con 13 grados de libertad resulte menor a él.
- Es decir: es el valor tal que hay un área de 0.1 a su izquierda, bajo la curva de la función de densidad de una variable chi-cuadrada con $v = 13$.
- Buscando en la tabla, vemos que vale: 4.107.

- Sea $X: \chi^2(50)$

- Se lee "X es una variable chi-cuadrada con 50 grados de libertad".

- El valor $\chi^2_{0.75;50}$:

- Es el fractil de la chi cuadrada para $\alpha = 0.75$ con 50 grados de libertad.
- Es decir: es el valor tal que hay probabilidad 0.75 de que una variable chi-cuadrada con 50 grados de libertad resulte menor a él.
- Es decir: es el valor tal que hay un área de 0.75 a su izquierda, bajo la curva de la función de densidad de una variable chi-cuadrada con $v = 50$.
- Buscando en la tabla, vemos que vale: 56.33.
- Como $v \geq 30$, también se puede calcular con la aproximación normal:

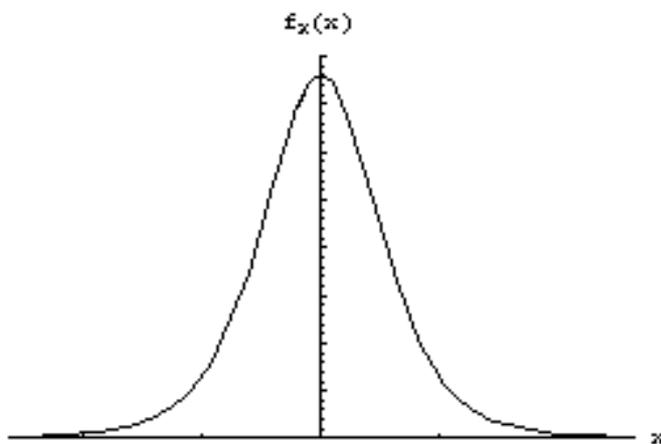
$$\chi^2_{0.75;50} = \frac{\left(z_{0.75} + \sqrt{2 \cdot 50 - 1} \right)}{2}$$

Buscando en la tabla de fractiles de la normal estándar, se obtiene el valor aproximado 56.44. Vemos que la aproximación es bastante buena.

Distribución t de Student

La variable aleatoria X tiene la distribución t de Student si su función de densidad de probabilidad es:

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right) \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\nu\pi}} \quad \forall x \in \mathfrak{R}$$



donde

- el parámetro ν es un número natural, y se conoce con el nombre de "grados de libertad"

- $$\Gamma(w) = \int_0^{+\infty} x^{w-1} e^{-x} dx$$

Notación

X:T(ν) se lee "X tiene una distribución t de Student con ν grados de libertad"

Propiedades

Si

$$\mathbf{X:T}(\nu)$$

entonces:

$$E(X) = \begin{cases} 0 & v \geq 2 \\ \frac{v}{2} & v = 1 \end{cases}$$

-

$$\sigma_x^2 = \begin{cases} v & v \geq 3 \\ \frac{v-2}{2} & v < 3 \end{cases}$$

-

Relación con otras distribuciones

Si una variable normal estándar se divide por la raíz de una variable chi-cuadrada dividida por sus grados de libertad, se obtiene una variable t de Student con la misma cantidad de grados de libertad que la chi-cuadrada. Es decir, si:

- $Z: N(0;1)$

- $X: \chi^2(v)$

$$Y = \frac{Z}{\sqrt{\frac{X}{v}}}$$

-

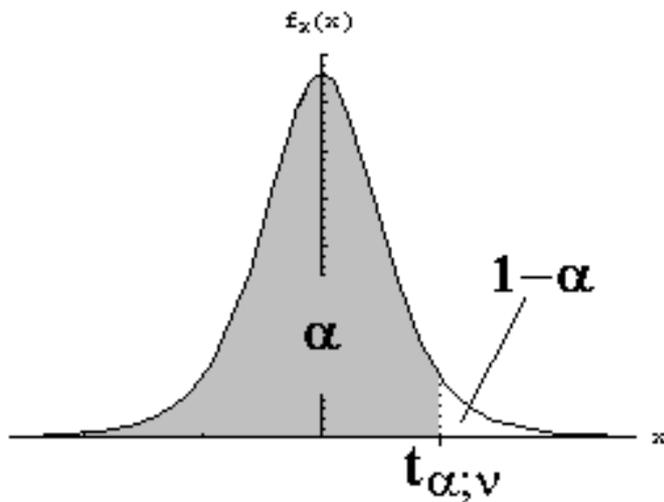
entonces:

$$Y: T(v)$$

Utilización

Debido al uso que le daremos (capítulo 8 en adelante), lo que nos interesa calcular de la distribución t de Student son sus fractiles. Es decir, los valores x tales que $P(X \leq x)$ es igual a un cierto α .

Sea $X: T(v)$, $t_{\alpha, v}$ es el valor x tal que $P(X \leq x) = \alpha$. Es decir, el valor tal que la probabilidad de que una variable t Student con v grados de libertad resulte menor que ese valor sea α . Dicho de otra forma, el valor que tiene un área α a la izquierda, bajo la curva de una t Student con v grados de libertad. Dichos fractiles se encuentran tabulados en el apéndice D.



Sobre esto, una advertencia: muchos autores trabajan con el fractil de la t de Student a derecha en vez de a izquierda, y por eso muchas propiedades y tablas tienen las convenciones contrarias a las que utilizaremos. Recomendamos al consultar una tabla verificar previamente si los fractiles son a izquierda o a derecha.

Si debiéramos usar una tabla con los fractiles a derecha para buscar $t_{\alpha;v}$, debemos buscar $t_{1-\alpha;v}$ ya que tener un área de α a la izquierda es lo mismo que tener un área de $1-\alpha$ a la derecha.

Otra aclaración: al igual que sucede con la distribución normal estándar, la distribución t de Student es simétrica respecto del cero. Luego, no vale la pena tener tabulados los fractiles para $\alpha > 0.5$ y $\alpha < 0.5$. En la tabla del apéndice D figuran solamente los fractiles con $\alpha > 0.5$. Los fractiles para $\alpha < 0.5$ pueden obtenerse con $t_{1-\alpha;v} = -t_{\alpha;v}$.

Aproximación

Para $v \geq 50$, se demuestra que la distribución t de Student con v grados de libertad es aproximadamente igual a la distribución normal estándar.

Es decir:

$$\left. \begin{array}{l} X : T(v) \\ v \geq 50 \end{array} \right\} \Rightarrow X : N(0; 1) \quad (\text{aprox.})$$

Consecuentemente, no vale la pena tener tabulados los fractiles de la t de Student para $v \geq 50$, porque las probabilidades se pueden obtener con:

$$P(X \leq x) \approx \Phi(x)$$

como si X fuera una normal estándar, y los fractiles se pueden obtener con:

$$t_{\alpha;v} \approx z_{\alpha}$$

donde z_{α} es el fractil de la distribución normal estándar con un área de α a la izquierda.

Ejemplos

- Sea $X:T(13)$
 - Se lee "X es una variable t de Student con 13 grados de libertad".
 - El valor $t_{0.9;13}$:
 - Es el fractil de la t de Student para $\alpha = 0.9$ con 13 grados de libertad.
 - Es decir: es el valor tal que hay probabilidad 0.9 de que una variable t de Student con 13 grados de libertad resulte menor a él.
 - Es decir: es el valor tal que hay un área de 0.9 a su izquierda, bajo la curva de la función de densidad de una variable t de Student con $v = 13$.
 - Buscando en la tabla, vemos que vale: 1.3502.
 - El valor $t_{0.1;13}$:
 - Es igual al anterior pero con $\alpha = 0.1$.
 - Es el fractil de la t de Student para $\alpha = 0.1$ con 13 grados de libertad.
 - Es decir: es el valor tal que hay probabilidad 0.1 de que una variable t de Student con 13 grados de libertad resulte menor a él.
 - Es decir: es el valor tal que hay un área de 0.1 a su izquierda, bajo la curva de la función de densidad de una variable t de Student con $v = 13$.
 - En la tabla no lo vamos a encontrar, porque la tabla tiene solamente los fractiles para $\alpha > 0.5$. Pero por la propiedad que vimos, $t_{1-\alpha;v} = -t_{\alpha;v}$, es decir, $t_{0.1;13} = -t_{0.9;13}$. En la tabla encontramos que $t_{0.9;13} = 1.3502$. Luego $t_{0.1;13} = -1.3502$. Esto es lógico porque la distribución t de Student es simétrica respecto del cero.
- Sea $X:T(50)$
 - Se lee "X es una variable t de Student con 50 grados de libertad".
 - El valor $t_{0.75;50}$:
 - Es el fractil de la t de Student para $\alpha = 0.75$ con 50 grados de libertad.
 - Es decir: es el valor tal que hay probabilidad 0.75 de que una variable t de Student con 50 grados de libertad resulte menor a él.
 - Es decir: es el valor tal que hay un área de 0.75 a su izquierda, bajo la curva de la función de densidad de una variable t de Student con $v = 50$.

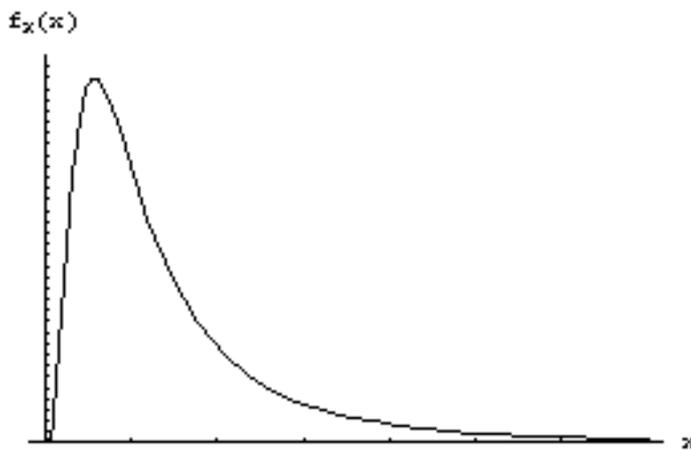
- Buscando en la tabla, vemos que vale: 0.6794.
- Como $v \geq 50$, también se puede calcular con la aproximación normal, es decir, $t_{\alpha, v} \approx z_{\alpha}$. Buscamos $z_{0.75}$ en la tabla y obtenemos 0.6745, valor bastante parecido a lo que obtuvimos sin aproximar.

Distribución F

La variable aleatoria X tiene la distribución F si su función de densidad de probabilidad es:

$$f_X(x) = \begin{cases} \frac{\Gamma\left(\frac{v_1 + v_2}{2}\right) \left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}} x^{\frac{v_1}{2}-1}}{\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right) \left(1 + \frac{xv_1}{v_2}\right)^{\frac{v_1 + v_2}{2}}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

donde



- los parámetros v_1 y v_2 son números naturales.
- $\Gamma(w) = \int_0^{+\infty} x^{w-1} e^{-x} dx$

Notación

$X:F(v_1, v_2)$ se lee "X tiene una distribución F con parámetros v_1 y v_2 "

Relación con otras distribuciones

Si se hace la división entre dos variables chi-cuadradas, cada una dividida previamente por sus grados de libertad, se obtiene una variable F, cuyos parámetros son los grados de libertad de las variables chi-cuadradas que estamos dividiendo. Los grados de libertad de la variable chi-cuadrado del numerador será el parámetro

v_1 , y los grados de libertad de la variable chi-cuadrado del denominador será el parámetro v_2 . Es decir, si:

- $X: \chi^2(v_1)$

- $Y: \chi^2(v_2)$

$$Z = \frac{X/v_1}{Y/v_2} = \frac{X v_2}{Y v_1}$$

-

entonces:

$$Z: F(v_1, v_2)$$

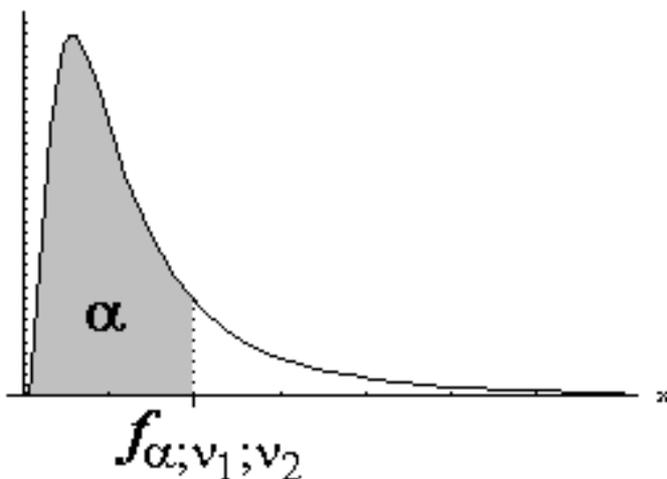
Utilización

Debido al uso que le daremos (capítulo 10), lo que nos interesa calcular de la distribución F son sus fractiles. Es decir, los valores x tales que $P(X \leq x)$ es igual a un cierto α .

Sea $X: F(v_1, v_2)$, $f_{\alpha; v_1, v_2}$ es el valor x tal que $P(X \leq x) = \alpha$. Es decir, el valor tal que la probabilidad de que una variable F con parámetros v_1 y v_2 resulte menor que ese valor sea α . Dicho de otra forma, el valor que tiene un área α a la izquierda, bajo la curva de una F con parámetros v_1 y v_2 .

Dichos fractiles se encuentran tabulados en el apéndice D.

$f_{\alpha}(x)$



Sobre esto, una advertencia: muchos autores trabajan con el fractil de la F a derecha en vez de a izquierda, y por eso muchas propiedades y tablas tienen las convenciones contrarias a las que utilizaremos. Recomendamos al consultar una tabla verificar previamente si los fractiles son a izquierda o a derecha. Si debiéramos

usar una tabla con los fractiles a derecha para buscar $f_{\alpha; v_1, v_2}$, debemos buscar

$f_{1-\alpha; v_1, v_2}$ ya que tener un área de α a la izquierda es lo mismo que tener un área de

$1-\alpha$ a la derecha.

Una propiedad importante a tener en cuenta es:

$$f_{\alpha;v_1,v_2} = \frac{1}{f_{1-\alpha;v_2,v_1}}$$

Es decir, el fractil de área α de una F con parámetros v_1 y v_2 , es uno sobre el fractil de área $1-\alpha$ de una F con parámetros v_2 y v_1 (es decir, intercambiados).

Luego, no vale la pena tener tabulados los fractiles para $\alpha > 0.5$ y $\alpha < 0.5$. En la tabla del apéndice D figuran solamente los fractiles con $\alpha > 0.5$. Los fractiles para $\alpha < 0.5$ pueden obtenerse con la propiedad enunciada.

Ejemplos

• Sea $X:F(5;10)$

• Se lee "X es una variable F con parámetros $v_1 = 5$ y $v_2 = 10$ ".

• El valor $f_{0.9;5;10}$:

• Es el fractil de la F para $\alpha = 0.9$ con parámetros $v_1 = 5$ y $v_2 = 10$.

• Es decir: es el valor tal que hay probabilidad 0.9 de que una variable F con parámetros $v_1 = 5$ y $v_2 = 10$ resulte menor a él.

• Es decir: es el valor tal que hay un área de 0.9 a su izquierda, bajo la curva de la función de densidad de una variable F con $v_1 = 5$ y $v_2 = 10$.

• Buscando en la tabla, vemos que vale: 2.52.

• Sea $X:F(15;8)$

• Se lee "X es una variable F con parámetros $v_1 = 15$ y $v_2 = 8$ ".

• El valor $f_{0.1;15;8}$:

• Es el fractil de la F para $\alpha = 0.1$ con parámetros $v_1 = 15$ y $v_2 = 8$.

• Es decir: es el valor tal que hay probabilidad 0.1 de que una variable F con parámetros $v_1 = 15$ y $v_2 = 8$ resulte menor a él.

• Es decir: es el valor tal que hay un área de 0.1 a su izquierda, bajo la curva de la función de densidad de una variable F con $v_1 = 15$ y $v_2 = 8$.

• En la tabla no lo encontraremos, porque $\alpha < 0.5$. Pero podemos usar la propiedad enunciada anteriormente,

$$f_{\alpha;v_1,v_2} = \frac{1}{f_{1-\alpha;v_2,v_1}}$$

es decir:

$$f_{0.1;15;8} = \frac{1}{f_{0.9;8;15}}$$

De la tabla obtenemos que $f_{0.9;8;15} = 2.12$.

Luego, $f_{0.1;15;8} = 0.47$

Distribución Beta

La variable aleatoria X tiene la distribución beta si su función de densidad de probabilidad es:

$$f_X(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} & 0 < x < 1 \\ 0 & \forall \text{ otro } x \end{cases}$$

donde

- los parámetros a y b son números reales positivos.

- $\Gamma(w) = \int_0^{+\infty} x^{w-1} e^{-x} dx$

Notación

$X:\text{Beta}(a,b)$ se lee "X tiene una distribución beta con parámetros a y b ".

Propiedades

Si

$X:\text{Beta}(a,b)$

entonces:

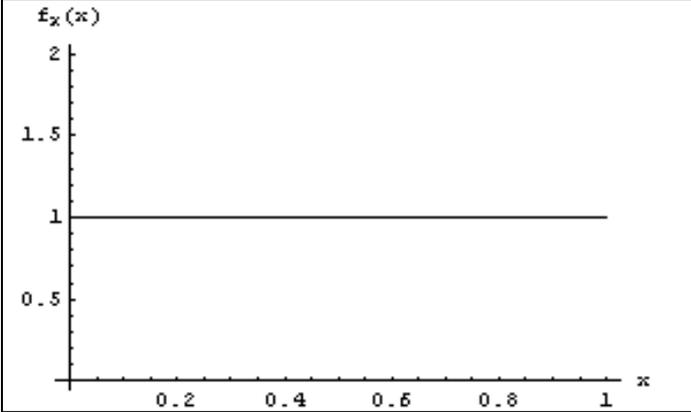
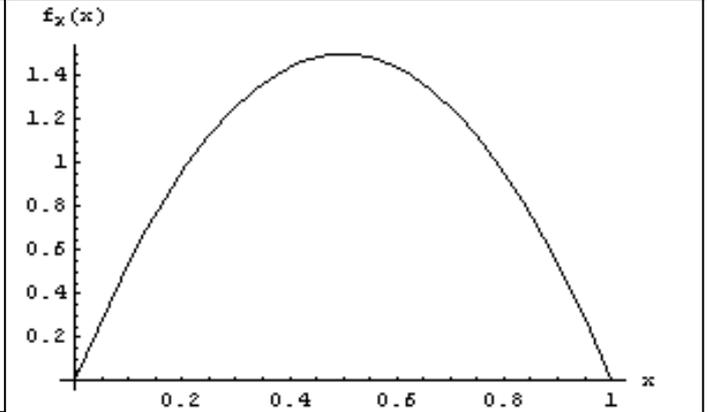
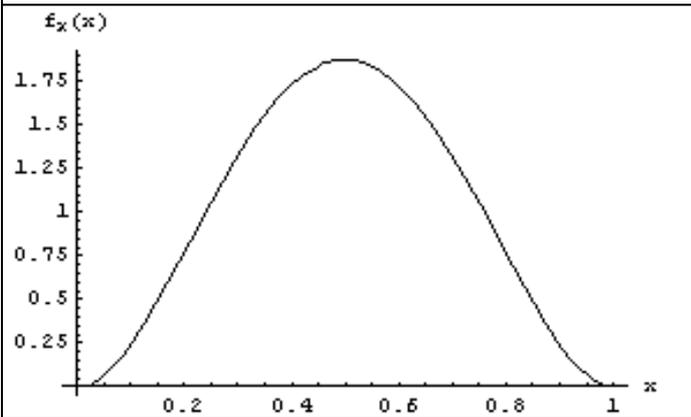
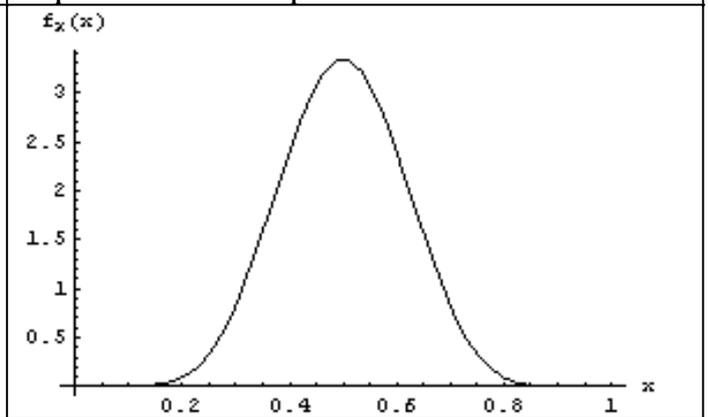
- $E(X) = \frac{a}{a+b}$
- $Mo(X) = \frac{a-1}{a+b-2}$
- $\sigma_x^2 = \frac{a b}{(a+b)^2 (a+b+1)}$

Aspecto

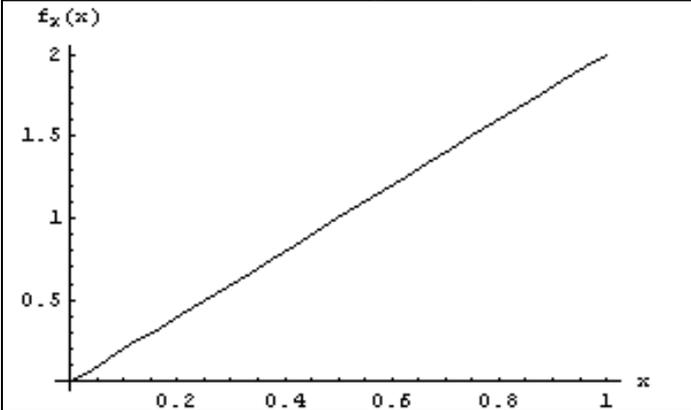
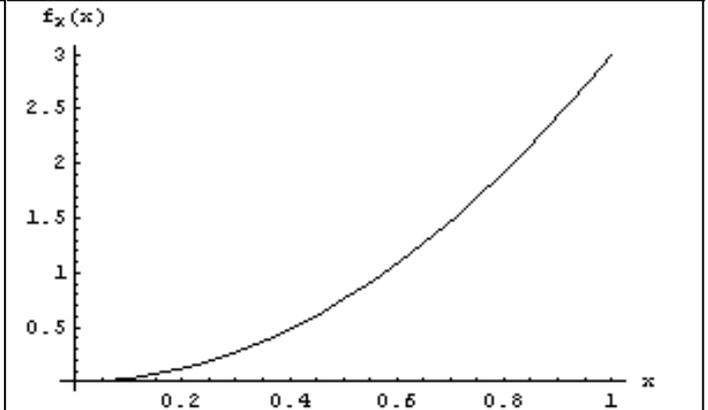
Una de las razones por las cuales nos será útil esta distribución (capítulo 11) es la diversidad de formas que puede adquirir, dependiendo del valor de los parámetros. A continuación graficaremos la distribución para distintos valores de a y b con el fin de mostrar la variedad de formas disímiles que esta distribución puede tomar. Comencemos por observar que:

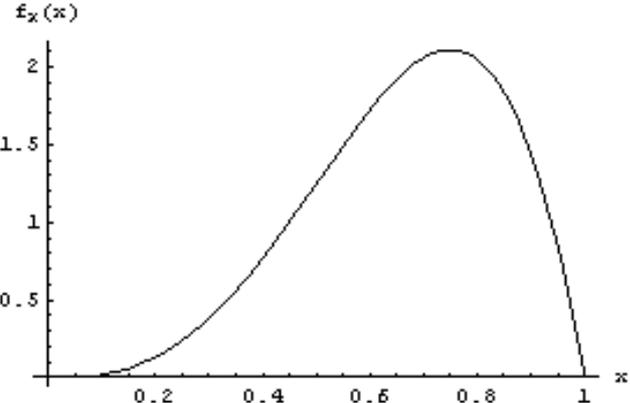
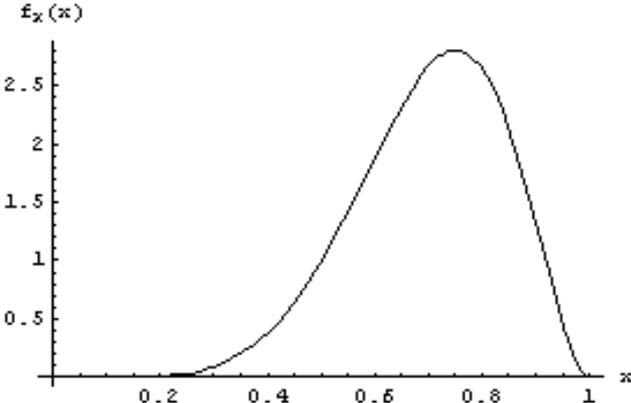
- La distribución siempre es no nula en el intervalo (0;1) y nula fuera de él.
- Los parámetros son números reales positivos, pero por razones que se estudiarán más adelante, solamente nos interesa darles valores naturales.

Veamos las formas que adquiere la distribución cuando $a = b$:

	
<p style="text-align: center;">a = 1 ; b = 1</p> <p>Cuando $a = b = 1$, la distribución beta coincide con la uniforme.</p>	<p style="text-align: center;">a = 2 ; b = 2</p> <p>Al valer 2 los parámetros, la distribución adquiere forma de parábola.</p>
	
<p style="text-align: center;">a = 3 ; b = 3</p> <p>A partir del valor 3, la distribución adquiere una forma acampanada.</p>	<p style="text-align: center;">a = 9 ; b = 9</p> <p>Del 3 en adelante, la distribución sigue siendo una campana, pero se va concentrando cada vez más.</p>

Ahora veamos la forma que adquiere cuando los parámetros son distintos:

	
---	--

<p style="text-align: center;">a = 2 ; b = 1</p> <p>Cuando un parámetro vale 2 y el otro vale 1, la distribución tiene forma de recta.</p>	<p style="text-align: center;">a = 3 ; b = 1</p> <p>Si un parámetro vale 1 y el otro vale 3 ó más, la distribución se va curvando.</p>
	
<p style="text-align: center;">a = 4 ; b = 2</p> <p>Si el parámetro menor vale 2, la distribución adquiere forma de campana inclinada hacia la izquierda si el parámetro que vale 2 es a, o hacia la derecha si el parámetro que vale 2 es b.</p>	<p style="text-align: center;">a = 7 ; b = 3</p> <p>En el caso general, si el menor parámetro es mayor o igual a 2, la distribución tiene forma de campana inclinada hacia la izquierda si el parámetro menor es a, y hacia la derecha si el parámetro menor es b.</p>

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 6 de junio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 12 de julio de 2004

CAPÍTULO VIII

Estimadores

Hasta ahora, los parámetros de las distribuciones con las que trabajamos eran dato, o se podían calcular a partir de otras distribuciones cuyos parámetros eran dato, o bien se podían calcular de forma evidente.

Pero en la vida real los valores de los parámetros no se conocen mágicamente, y generalmente se obtienen haciendo experimentos para poder estimarlos.

Entonces si por ejemplo sabemos que la longitud de los tornillos que produce una determinada máquina es una variable normal, pero no sabemos cuánto vale el parámetro μ de esa distribución normal, podemos hacer el experimento de tomar 10 tornillos, calcular el promedio de sus longitudes, y usar ese promedio como estimación de μ .

Un estimador de un determinado parámetro es una cuenta que se hace con los resultados de un experimento para estimar el valor del parámetro. En el ejemplo anterior, el parámetro es μ , y en este caso el estimador es:

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}$$

A partir de ahora usaremos algunas notaciones y convenciones nuevas:

- Llamaremos muestra al conjunto de los n resultados que obtenemos al hacer un experimento n veces. Por ejemplo, si para estudiar el peso de las manzanas en un lote de un millón de manzanas tomamos 20 y las pesamos, el experimento es "tomar una manzana del lote y pesarla", n vale 20, y la muestra son los 20 valores obtenidos. El valor de n se conoce como "tamaño de la muestra".
- Llamaremos población al conjunto de infinitos elementos de donde extraemos la muestra. Dicho número puede no ser estrictamente infinito, pero asumiremos que lo es.
- Llamaremos x_i a cada resultado del experimento (el peso de cada una de las 20 manzanas observadas).
- Notaremos \bar{X} al promedio de los valores de la muestra. Es decir,

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

- θ es un parámetro que queremos estimar, por ejemplo, la media μ de la distribución normal de pesos de las manzanas.
- $\hat{\theta}$ es un estimador que usaremos para estimar el valor de θ .
- Nunca olvidar que $\theta \neq \hat{\theta}$

Dijimos que un estimador es simplemente una cuenta que se hace con los resultados del experimento para estimar el parámetro, con lo cual en esencia cualquier función de dichos resultados puede usarse como estimador de un determinado parámetro.

Es importante entonces determinar:

- a) qué propiedades debe tener un estimador para ser considerado "bueno".
- b) qué importancia relativa tiene cada una de esas propiedades. Dicho de otra forma: cómo determinar, dados dos estimadores, cuál es el mejor.

Antes de dedicarnos a dicha cuestión, haremos algunas consideraciones sobre el valor esperado de los estimadores:

Valor esperado de un estimador

Llamamos $E(\hat{\theta})$ al valor esperado del estimador $\hat{\theta}$, en términos del parámetro θ . Notemos que no es necesario conocer el verdadero valor del parámetro θ para calcular el valor esperado del estimador en términos de θ .

Por ejemplo, si θ es la media desconocida de una población normal, y para estimarla se toma una muestra de tamaño 2 cuyos valores son x_1 y x_2 , y el estimador utilizado es

$$\hat{\theta} = \frac{3}{4}x_1 + \frac{1}{2}x_2$$

entonces:

$$E(\hat{\theta}) = E\left(\frac{3}{4}x_1 + \frac{1}{2}x_2\right) = \frac{3}{4}E(x_1) + \frac{1}{2}E(x_2)$$

y como x_1 y x_2 fueron tomados de una población cuya media es justamente θ , entonces:

$$\frac{3}{4}E(x_1) + \frac{1}{2}E(x_2) = \frac{3}{4}\theta + \frac{1}{2}\theta = \frac{5}{4}\theta$$

Con lo cual vemos que aunque en ningún momento supimos cuánto vale θ , podemos determinar que el valor esperado del estimador $\hat{\theta}$ es $\frac{5}{4}\theta$

Propiedades deseables de un estimador

Un estimador $\hat{\theta}$ del parámetro θ es:

- Insesgado, si $E(\hat{\theta}) = \theta$

Es decir, si el valor esperado del estimador es el parámetro. Aunque no conozcamos el valor del parámetro, podemos igualmente calcular la esperanza del estimador en términos del parámetro y ver si coinciden. Evidentemente es muy deseable que un estimador sea insesgado.

- Consistente, si $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$

Es decir, si cuando el tamaño de la muestra tiende a infinito, la esperanza del estimador tiende al parámetro estimado.

Notemos que si un estimador es insesgado, necesariamente también es consistente. Es deseable que un estimador sea consistente, porque en ese caso, si la muestra es lo suficientemente grande, el estimador tendrá propiedades similares a las de los estimadores insesgados.

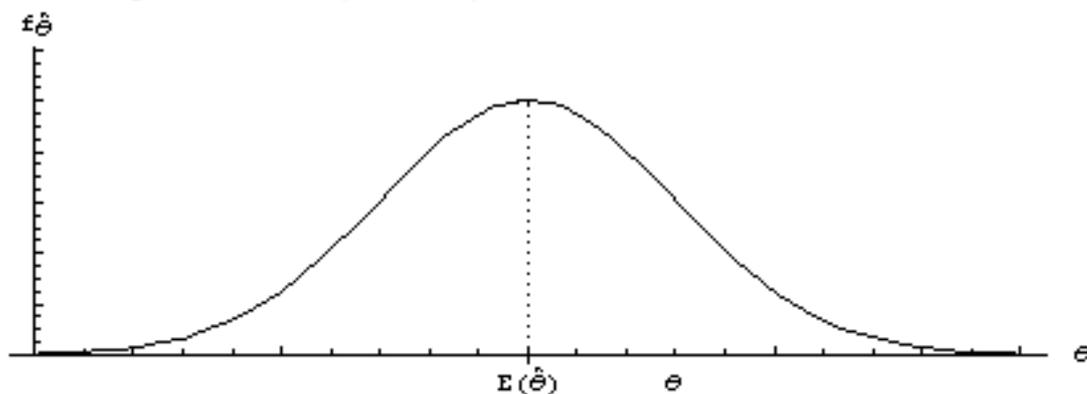
- Eficiente, si $\lim_{n \rightarrow \infty} \sigma^2(\hat{\theta}) = 0$

Es decir, si cuando el tamaño de la muestra tiende a infinito, la varianza del estimador tiende a cero.

Es deseable que un estimador sea eficiente, porque en dicho caso, si la muestra es lo suficientemente grande, la variabilidad del estimador será pequeña y entonces su valor estará cerca de su valor esperado (el cual, si el estimador es consistente, estará cerca del parámetro).

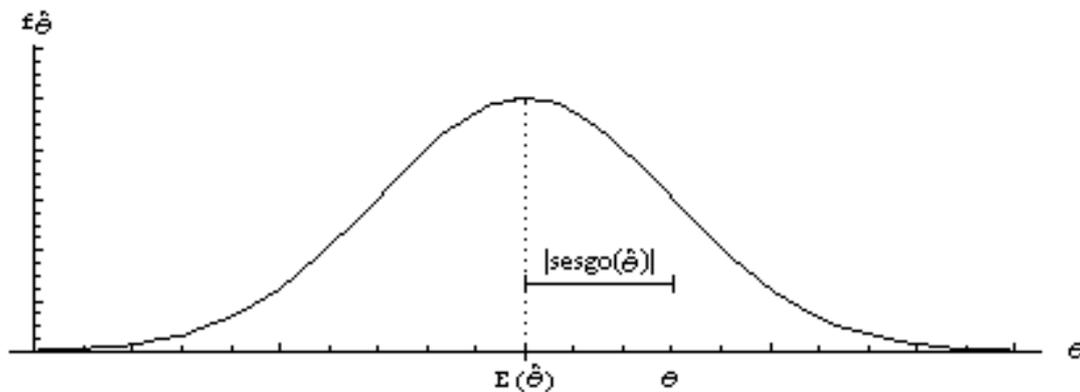
Veamos una interpretación gráfica de estas propiedades deseables. Supongamos que estamos estimando la media de una población normal. Es decir, la media de una población que sabemos que es normal, aunque no sepamos su media. Si como estimador de la media usamos, por ejemplo, alguna combinación lineal de los valores de una muestra tomada de esa población, entonces como el valor de cada valor de la muestra es una variable normal en si misma, y una combinación lineal de variables normales es una variable normal, nuestro estimador también es una variable aleatoria normal.

Si calculáramos como vimos antes el valor esperado del estimador y lo graficáramos, podríamos llegar a un gráfico como este:

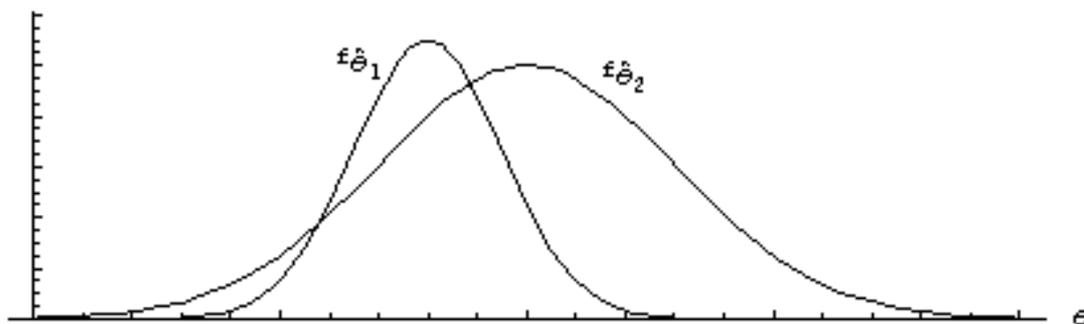


En este gráfico podemos apreciar que es deseable que el valor esperado del estimador coincida con el parámetro estimado. Denominamos sesgo a la diferencia

$E(\hat{\theta}) - \theta$. Por eso cuando el sesgo de un estimador es cero, se lo denomina "insesgado".



Como podemos observar, el estimador graficado no es insesgado. Por lo que dijimos antes, es deseable que el sesgo de un estimador sea pequeño. Otra característica importante que analizamos fue la varianza. Es deseable que la varianza de un estimador sea pequeña, para que la variabilidad respecto de su valor esperado sea pequeña.



En el ejemplo graficado, la varianza de $\hat{\theta}_1$ es más pequeña que la de $\hat{\theta}_2$. Vemos que su variabilidad respecto de su valor esperado es menor.

Comparación de estimadores

El error cuadrático medio de un estimador se define:

$$W_{\hat{\theta}}^2 = \sigma_{\hat{\theta}}^2 + (E(\hat{\theta}) - \theta)^2$$

donde $\hat{\theta}$ es el estimador y θ el parámetro estimado.

Entre dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ el mejor será el que tenga el menor error cuadrático medio.

Observamos que el error cuadrático medio es la suma de la varianza y el cuadrado del sesgo. Podemos verificar que esto efectivamente refleja que a mayor varianza y a mayor sesgo, peor resulta el estimador.

Si los estimadores no son insesgados, el error cuadrático medio quedará en función del parámetro, en cuyo caso puede suceder que la decisión de cuál estimador es mejor dependa del valor del parámetro. En ese caso, es necesario dejar expresada

una regla de decisión basada en una inecuación, que para un determinado valor del parámetro permita decidir cuál de los dos estimadores es mejor. Dicho caso se analiza en el problema N°2.

Problemas típicos

1) Se sabe que una determinada población tiene una distribución normal, y que su desvío estándar es 1, pero se desconoce el valor de la media. Para estimar la media μ , se toma una muestra de tamaño 2 de la población, y se proponen 3 estimadores:

$$\hat{\mu}_1 = \frac{3}{4}x_1 + \frac{1}{2}x_2 \quad \hat{\mu}_2 = \frac{3}{4}x_1 + \frac{1}{4}x_2 \quad \hat{\mu}_3 = \frac{1}{2}x_1 + \frac{1}{2}x_2$$

Indique cuál es los 3 es el mejor.

Resolución

Vamos a calcular el error cuadrático medio de cada estimador y ver cuál de los estimadores posee el menor.

$$W_{\theta}^2 = \sigma_{\theta}^2 + (E(\hat{\theta}) - \theta)^2$$

Comenzamos entonces por calcular las esperanzas.

$$E(\hat{\mu}_1) = E\left(\frac{3}{4}x_1 + \frac{1}{2}x_2\right) = \frac{3}{4}E(x_1) + \frac{1}{2}E(x_2)$$

Como x_1 y x_2 fueron tomados de una determinada población, entonces sus valores son en esencia variables aleatorias, cuya distribución es la distribución de la población que estamos estudiando. El valor esperado de la población que estamos estudiando es μ . Entonces el valor esperado de x_1 y x_2 es el μ que queremos estimar, con lo cual:

$$\frac{3}{4}E(x_1) + \frac{1}{2}E(x_2) = \frac{3}{4}\mu + \frac{1}{2}\mu = \frac{5}{4}\mu$$

$$E(\hat{\mu}_1) = \frac{5}{4}\mu$$

y obtenemos que

Análogamente:

$$E(\hat{\mu}_2) = E\left(\frac{3}{4}x_1 + \frac{1}{4}x_2\right) = \frac{3}{4}E(x_1) + \frac{1}{4}E(x_2) = \frac{3}{4}\mu + \frac{1}{4}\mu = \mu$$

$$E(\hat{\mu}_3) = E\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) = \frac{1}{2}E(x_1) + \frac{1}{2}E(x_2) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

Vemos entonces que $\hat{\mu}_1$ es sesgado, mientras que $\hat{\mu}_2$ y $\hat{\mu}_3$ son insesgados.

Calculemos las varianzas:

$$\sigma^2(\hat{\mu}_1) = \sigma^2\left(\frac{3}{4}x_1 + \frac{1}{2}x_2\right)$$

Como vale $\sigma^2(aX + bY) = a^2\sigma_x^2 + b^2\sigma_y^2$ entonces queda:

$$\sigma^2(\hat{\mu}_1) = \sigma^2\left(\frac{3}{4}x_1 + \frac{1}{2}x_2\right) = \frac{9}{16}\sigma_{x_1}^2 + \frac{1}{4}\sigma_{x_2}^2$$

Como x_1 y x_2 siguen la distribución de la población, cuya varianza es conocida y vale 1, entonces queda:

$$\sigma^2(\hat{\mu}_1) = \frac{9}{16}\sigma_{x_1}^2 + \frac{1}{4}\sigma_{x_2}^2 = \frac{9}{16} + \frac{1}{4} = 0.8125$$

Análogamente:

$$\sigma^2(\hat{\mu}_2) = \sigma^2\left(\frac{3}{4}x_1 + \frac{1}{4}x_2\right) = \frac{9}{16}\sigma_{x_1}^2 + \frac{1}{16}\sigma_{x_2}^2 = \frac{9}{16} + \frac{1}{16} = 0.625$$

$$\sigma^2(\hat{\mu}_3) = \sigma^2\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) = \frac{1}{4}\sigma_{x_1}^2 + \frac{1}{4}\sigma_{x_2}^2 = \frac{1}{4} + \frac{1}{4} = 0.5$$

Con lo cual ya estamos en condiciones de calcular los errores cuadráticos medios:

$$W_{\hat{\mu}_1}^2 = \sigma_{\hat{\mu}_1}^2 + (E(\hat{\mu}_1) - \mu)^2 = 0.8125 + \frac{1}{16}\mu^2$$

$$W_{\hat{\mu}_2}^2 = \sigma_{\hat{\mu}_2}^2 + (E(\hat{\mu}_2) - \mu)^2 = 0.625$$

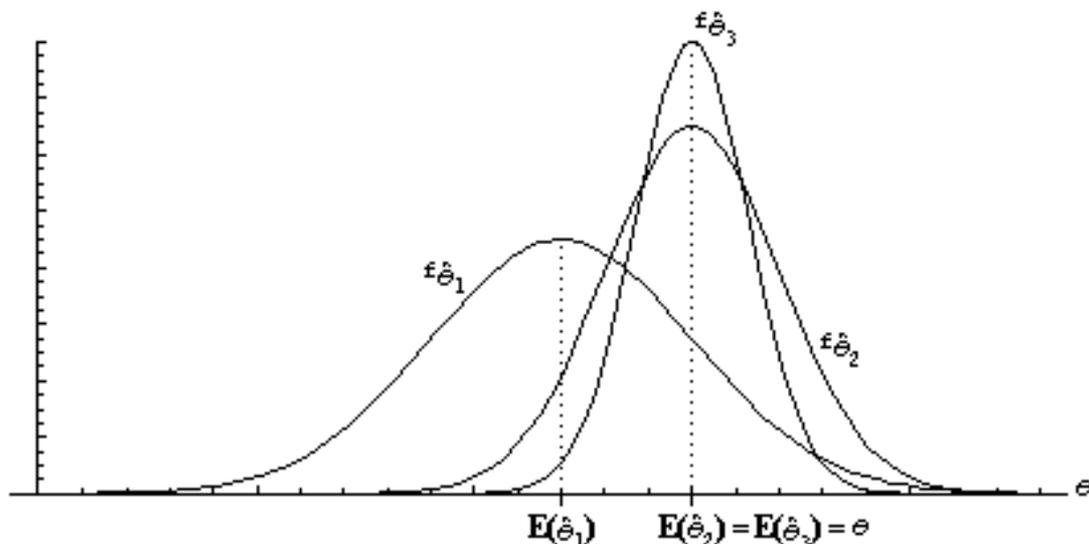
$$W_{\hat{\mu}_3}^2 = \sigma_{\hat{\mu}_3}^2 + (E(\hat{\mu}_3) - \mu)^2 = 0.5$$

Vemos que como el segundo término de $W_{\hat{\mu}_1}^2$ siempre es positivo, entonces $W_{\hat{\mu}_1}^2 \geq 0.8125$, con lo cual podremos ordenar, sin ninguna duda:

$$W_{\hat{\mu}_3}^2 < W_{\hat{\mu}_2}^2 < W_{\hat{\mu}_1}^2$$

de donde resulta que el mejor estimador es $\hat{\mu}_3$.

A modo ilustrativo, podríamos hacer el siguiente gráfico cualitativo:



2) Se desea estimar el parámetro μ de una determinada población Poisson, para lo cual se extrae una muestra de tamaño 3 y se proponen 2 estimadores:

$$\hat{\mu}_1 = \frac{x_1 + x_2 + x_3}{5}$$

$$\hat{\mu}_2 = \frac{2x_1 + 3x_2 + 5x_3}{10}$$

Indique cuál es preferible.

Resolución

Vamos a calcular el error cuadrático medio de cada estimador y ver cuál de los estimadores posee el menor.

Comenzamos entonces por calcular las esperanzas.

$$E(\hat{\mu}_1) = E\left(\frac{x_1 + x_2 + x_3}{5}\right) = \frac{1}{5} E(x_1 + x_2 + x_3) = \frac{1}{5} (E(x_1) + E(x_2) + E(x_3))$$

Como x_1 , x_2 , y x_3 fueron tomados de una determinada población, entonces sus valores son en esencia variables aleatorias, cuya distribución es la distribución de la población que estamos estudiando. El valor esperado de la población que estamos estudiando es μ . Entonces el valor esperado de x_1 , x_2 , y x_3 es el μ que queremos estimar, con lo cual:

$$\frac{1}{5} (E(x_1) + E(x_2) + E(x_3)) = \frac{1}{5} (\mu + \mu + \mu) = \frac{3\mu}{5}$$

$$E(\hat{\mu}_1) = \frac{3}{5}\mu$$

y obtenemos que

Análogamente:

$$E(\hat{\mu}_2) = E\left(\frac{2x_1 + 3x_2 + 5x_3}{10}\right) = E\left(\frac{2}{10}x_1 + \frac{3}{10}x_2 + \frac{5}{10}x_3\right) = \frac{2}{10}E(x_1) + \frac{3}{10}E(x_2) + \frac{5}{10}E(x_3) = \frac{2}{10}\mu + \frac{3}{10}\mu + \frac{5}{10}\mu = \mu$$

Vemos entonces que $\hat{\mu}_1$ es sesgado, mientras que $\hat{\mu}_2$ es insesgado.

Calculemos las varianzas:

$$\sigma^2(\hat{\mu}_1) = \sigma^2\left(\frac{x_1 + x_2 + x_3}{5}\right) = \frac{1}{25}\sigma_{x_1}^2 + \frac{1}{25}\sigma_{x_2}^2 + \frac{1}{25}\sigma_{x_3}^2$$

Como x_1 , x_2 , y x_3 siguen la distribución de la población, y la varianza de una distribución Poisson con parámetro μ vale μ igual que la media, entonces queda:

$$\sigma^2(\hat{\mu}_1) = \frac{1}{25}\sigma_{x_1}^2 + \frac{1}{25}\sigma_{x_2}^2 + \frac{1}{25}\sigma_{x_3}^2 = \frac{3}{25}\mu = 0.12\mu$$

Análogamente:

$$\sigma^2(\hat{\mu}_2) = \sigma^2\left(\frac{2x_1 + 3x_2 + 5x_3}{10}\right) = \frac{1}{64}(4\sigma_{x_1}^2 + 9\sigma_{x_2}^2 + 25\sigma_{x_3}^2) = \frac{1}{64}38\mu = 0.59\mu$$

Con lo cual ya estamos en condiciones de calcular los errores cuadráticos medios:

$$W_{\hat{\mu}_1}^2 = 0.12\mu + \frac{4}{25}\mu^2$$

$$W_{\hat{\mu}_2}^2 = 0.59\mu$$

Vemos que los errores nos quedaron en función de μ , parámetro que justamente desconocemos.

Lo que podemos hacer en estos casos es enunciar una regla de decisión que indique para qué valores de μ conviene cada estimador.

Para que $W_{\hat{\mu}_1}^2$ sea mejor que $W_{\hat{\mu}_2}^2$ debe cumplirse:

$$W_{\hat{\mu}_1}^2 < W_{\hat{\mu}_2}^2$$

El resto del problema es simplemente hacer cuentas para encontrar los valores de μ

que satisfacen la inecuación

$$0.12\mu + \frac{4}{25}\mu^2 < 0.59\mu$$

$$\frac{4}{25}\mu^2 < 0.47\mu$$

Suponiendo $\mu \neq 0$ para poder simplificar, y considerando que μ no puede ser negativo porque si lo fuera no cumpliría con la inecuación que tenemos hasta ahora, queda:

$$\frac{4}{25}\mu < 0.47$$

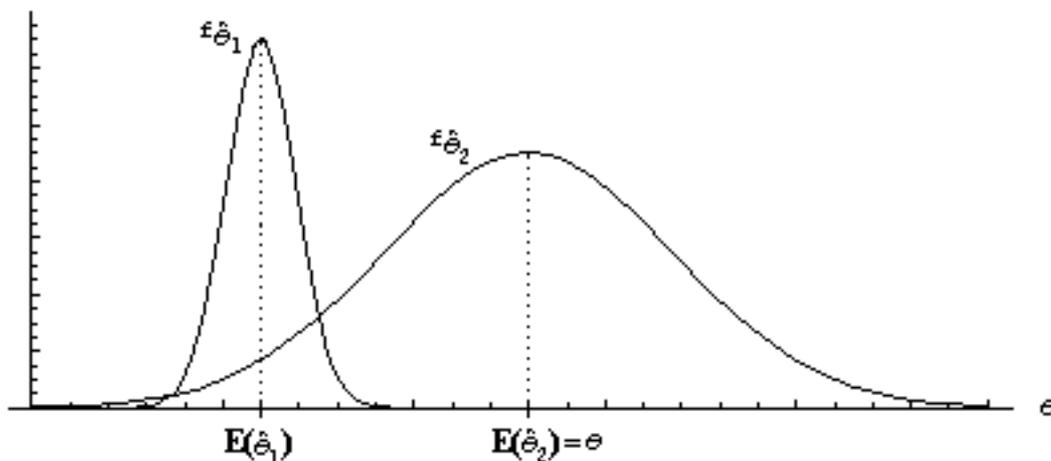
$$\mu < \frac{0.47 \cdot 25}{4}$$

$$\mu < 2.94$$

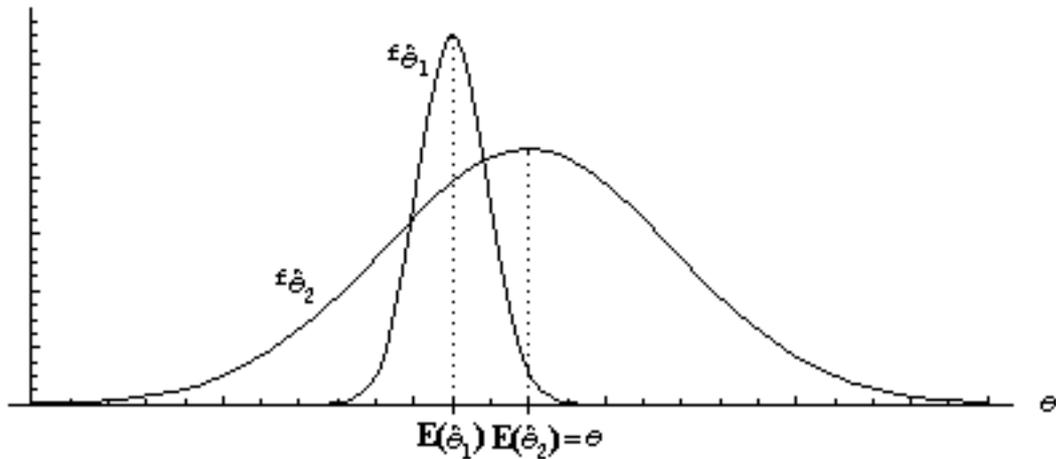
Con lo cual si μ es menor que 2.94 convendrá usar $\hat{\mu}_1$ y en caso contrario convendrá usar $\hat{\mu}_2$.

Podría caber la pregunta: ¿cómo sé si la inecuación se cumple si justamente no conozco el valor de μ ? La respuesta es, que como x_1 , x_2 , y x_3 son valores tomados de una población Poisson con media μ , mirándolos puedo tener una idea muy aproximada de cuánto vale μ . Incluso podría usar las fórmulas de los estimadores para estimar μ por ambos, y luego si los dos resultados satisfacen o niegan la inecuación, me quedaré con el que corresponda. Si por ejemplo estimo μ a partir de los datos mediante ambos estimadores, y según ambos μ tiene un valor muy superior a 2.94, entonces puedo determinar que $\hat{\mu}_2$ es mejor que $\hat{\mu}_1$ y entonces quedarme con el resultado de $\hat{\mu}_2$.

Por último, y aunque no lo piden, hagamos como ejemplo gráficos cualitativos de cómo serían los estimadores. En el caso de que μ sea grande ($\mu \gg 2.94$) el gráfico queda:



Vemos que con μ grande, el cuadrado del sesgo de $\hat{\mu}_1$ es grande y por eso $\hat{\mu}_2$ es mejor. Si μ fuera chico ($\mu < 2.94$), el gráfico quedaría así:



Podemos ver que si μ es chico, el cuadrado del sesgo de $\hat{\mu}_1$ es chico, y aunque sea sesgado, puede terminar siendo mejor estimador que $\hat{\mu}_2$ debido a que la varianza de este es más grande.

3) Se desea estimar el parámetro μ de una población normal $N(\mu ; \sigma)$. Se toma una muestra de tamaño n , y se propone como estimador el promedio muestral. Indique si el estimador propuesto es insesgado, consistente y eficiente, e indique qué conclusión se puede extraer.

Resolución

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} \quad (\text{promedio de la muestra})$$

Para que sea insesgado debe cumplirse $E(\hat{\mu}) = \mu$

$$E(\hat{\mu}) = E\left(\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{1}{n} E\left(\sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n E(x_i)$$

como los x_i vienen de una población normal, sus valores son variables aleatorias con la distribución de la población, con lo cual su valor esperado es el μ desconocido, y queda:

$$\frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu$$

con lo cual $E(\hat{\mu}) = \mu$

=> el estimador es insesgado.

Para que sea consistente debe cumplirse $\lim_{n \rightarrow \infty} E(\hat{\mu}) = \mu$

Como $E(\hat{\mu}) = \mu$, necesariamente también se cumple $\lim_{n \rightarrow \infty} E(\hat{\mu}) = \mu$
=> el estimador es consistente.

Para que sea eficiente debe cumplirse $\lim_{n \rightarrow \infty} \sigma^2(\hat{\mu}) = 0$

$$\sigma^2(\hat{\mu}) = \sigma^2 \left(\frac{\sum_{i=1}^n x_i}{n} \right) = \frac{1}{n^2} \sigma^2 \left(\sum_{i=1}^n x_i \right) = \frac{1}{n^2} \sum_{i=1}^n \sigma_{x_i}^2$$

El desvío del valor de cada x_i es el mismo desvío que el de la población de la cual provienen, con lo cual $\sigma^2(x_i) = \sigma^2$ y entonces:

$$\frac{1}{n^2} \sum_{i=1}^n \sigma_{x_i}^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$$

con lo cual $\sigma^2(\hat{\mu}) = \frac{\sigma^2}{n}$ y eso cumple con $\lim_{n \rightarrow \infty} \sigma^2(\hat{\mu}) = 0$

=> el estimador es eficiente.

La conclusión que se puede sacar es que el promedio de la muestra es un muy buen estimador para la media de una población normal.

4) Se sabe que sesgo del estimador $\hat{\theta}$ vale $\frac{3}{4}\theta$. Proponga un estimador insesgado para reemplazarlo.

Resolución

Si $\text{sesgo}(\hat{\theta}) = \frac{3}{4}\theta$ entonces $E(\hat{\theta}) - \theta = \frac{3}{4}\theta$, con lo cual $E(\hat{\theta}) = \frac{7}{4}\theta$.

Propondremos el estimador $\hat{\theta}_2 = \frac{4}{7}\hat{\theta}$, porque si $E(\hat{\theta}) = \frac{7}{4}\theta$ entonces:

$E(\hat{\theta}_2) = E\left(\frac{4}{7}\hat{\theta}\right) = \frac{4}{7}E(\hat{\theta}) = \frac{4}{7} \cdot \frac{7}{4}\theta = \theta$, con lo cual $\hat{\theta}_2$ resulta insesgado.

Estimadores de máxima verosimilitud

Ahora ya sabemos cuándo un estimador es bueno, o bien cuál entre dos estimadores es más conveniente. Pero nuestro estudio se basó en analizar determinados estimadores que nos fueron propuestos.

En ningún momento propusimos nosotros los estimadores.

Vimos que, por ejemplo, el promedio de la muestra resulta ser un muy buen estimador para la media de una población. Pero esto no es demasiado sorprendente. Es probablemente lo primero que usaríamos, aunque no supiéramos nada de lo estudiado en este capítulo.

¿Qué sucede si nos toca a nosotros proponer un estimador para un determinado parámetro de una determinada distribución? ¿Sería siempre tan obvio como en el caso de tomar el promedio muestral para aproximar la media poblacional? Es decir, el problema es que pueden darse parámetros de distribuciones para los cuales no resulta tan obvio cómo construir un buen estimador.

Para eso contamos con un método denominado "máxima verosimilitud". El mismo se basa en proponer la hipótesis de que sucede "lo que mayores probabilidades tenía de suceder". Es decir, que en cada valor obtenido en la muestra, se obtuvo el valor que más probabilidades tenía de obtenerse.

A tal efecto, se construye una función denominada "función de verosimilitud", que para una determinada muestra x_1, x_2, \dots, x_n devuelve la probabilidad de que se obtengan exactamente esos valores. Luego se busca el máximo de dicha función en términos del parámetro, con lo cual se encuentran los x_i para los cuales la probabilidad de que salgan es máxima. Por último se deja el parámetro en función de los resultados, con lo cual lo que se obtiene es un estimador del parámetro, es decir, una cuenta con los resultados que sirve para estimar el parámetro.

Los estimadores obtenidos por el método de máxima verosimilitud tienen propiedades deseables. No son necesariamente insesgados, pero son en general convergentes y eficientes.

A continuación veremos el método enunciado, luego un ejemplo, y luego la justificación del método.

Método

(implementación práctica del método de máxima verosimilitud)

1) Escribir la función de verosimilitud:

$$L = \prod_{i=1}^n P_X(x_i) \quad \text{si } X \text{ es discreta}$$

$$L = \prod_{i=1}^n f_X(x_i) \quad \text{si } X \text{ es continua}$$

2) Tomar $A = \ln(L)$ y transformar las productorias en sumatorias y las potencias en factores, y simplificar las sumatorias lo más posible.

- 3) Derivar A respecto del parámetro, e igualar la derivada a cero.
- 4) Intentar despejar el parámetro en función de las x_i . Si se puede, lo que se obtiene es el estimador, y el método termina.
- 5) Si no se puede, porque no depende de las x_i o bien porque depende pero no se puede hallar un valor no trivial del parámetro, ver más adelante en esta misma sección la explicación bajo el título "casos patológicos".

Ejemplo

Encontrar el estimador de máxima verosimilitud para el parámetro p de una binomial $X:Bi(m;p)$. La binomial se escribió con parámetro m para no confundirlo con el n del tamaño de la muestra.

$$P(X = x) = \binom{m}{x_i} p^x (1-p)^{m-x}$$

Vemos que P_x no es constante respecto de las x_i . Por lo tanto podemos aplicar el método descripto más arriba.

$$L = \prod_{i=1}^n P_x(x_i)$$

1)

2) Tomamos $A = \ln(L)$. Para simplificar A , recordemos las propiedades de los logaritmos:

- $\ln(a \cdot b) = \ln(a) + \ln(b)$
- $\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b)$
- $\ln(a^b) = b \ln(a)$

Aplicándolas podemos hacer:

$$\begin{aligned} A = \ln(L) &= \ln\left(\prod_{i=1}^n P_x(x_i)\right) = \sum_{i=1}^n \ln(P_x(x_i)) = \sum_{i=1}^n \ln\left[\binom{m}{x_i} p^{x_i} (1-p)^{m-x_i}\right] = \\ &= \sum_{i=1}^n \left[\ln\left(\binom{m}{x_i}\right) + \ln(p^{x_i}) + \ln[(1-p)^{m-x_i}] \right] = \sum_{i=1}^n \ln\left(\binom{m}{x_i}\right) + \sum_{i=1}^n \ln(p^{x_i}) + \sum_{i=1}^n \ln[(1-p)^{m-x_i}] = \\ &= \sum_{i=1}^n \ln\left(\binom{m}{x_i}\right) + \sum_{i=1}^n x_i \ln(p) + \sum_{i=1}^n (m-x_i) \ln[1-p] = \sum_{i=1}^n \ln\left(\binom{m}{x_i}\right) + \ln(p) \sum_{i=1}^n x_i + n m \ln[1-p] - \ln[1-p] \sum_{i=1}^n x_i \end{aligned}$$

con lo cual hemos transformado las productorias en sumatorias y las potencias en factores, y simplificamos las sumatorias y en general todo lo más posible, teniendo en cuenta que vamos a querer derivar.

3) Ahora derivamos A respecto del parámetro (en este caso, p).

$$\frac{dA}{dp} = 0 + \frac{\sum_{i=1}^n x_i}{p} + \frac{n m (-1)}{1-p} - \frac{(-1) \sum_{i=1}^n x_i}{1-p} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n m}{1-p} + \frac{\sum_{i=1}^n x_i}{1-p}$$

Igualamos la derivada a cero:

$$\frac{dA}{dp} = \frac{\sum_{i=1}^n x_i}{p} - \frac{nm}{1-p} + \frac{\sum_{i=1}^n x_i}{1-p} = 0$$

4) Despejamos el parámetro:

$$\frac{\sum_{i=1}^n x_i}{p} - \frac{nm}{1-p} + \frac{\sum_{i=1}^n x_i}{1-p} = 0 \Rightarrow \frac{\sum_{i=1}^n x_i}{p} = \frac{nm}{1-p} - \frac{\sum_{i=1}^n x_i}{1-p} \Rightarrow \frac{\sum_{i=1}^n x_i}{p} = \frac{nm - \sum_{i=1}^n x_i}{1-p} \Rightarrow$$

$$\Rightarrow \frac{1-p}{p} = \frac{nm - \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i} \Rightarrow \frac{1-p}{p} = \frac{nm}{\sum_{i=1}^n x_i} - \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i} \Rightarrow \frac{1}{p} - 1 = \frac{nm}{\sum_{i=1}^n x_i} - 1 \Rightarrow$$

$$\Rightarrow \frac{1}{p} = \frac{nm}{\sum_{i=1}^n x_i} \Rightarrow p = \frac{\sum_{i=1}^n x_i}{nm} = \frac{X}{m}$$

$$\hat{p} = \frac{X}{m}$$

Con lo cual obtuvimos que $\hat{p} = \frac{X}{m}$ es el estimador de máxima verosimilitud para el parámetro p de una distribución binomial $Bi(m;p)$.

Justificación del método

Se recomienda ir repasando el ejemplo dado a medida que se lee esta justificación. Comencemos por indicar que se asume que los n valores x_i de la muestra son todos independientes.

Luego, como L es la productoria de las $P_{x_i}(x_i)$ ó $f_{x_i}(x_i)$ evaluadas en cada uno de los x_i , L es la probabilidad de que ocurra simultáneamente $X_1 = x_1 \wedge X_2 = x_2 \wedge \dots \wedge X_n = x_n$. (Recordemos que si dos sucesos A y B son independientes, entonces $P(A \cap B) = P(A)P(B)$).

Si θ es el parámetro que queremos estimar, entonces las $P_{x_i}(x_i)$ ó $f_{x_i}(x_i)$ dependerán de él.

Consecuentemente, como L es una productoria de esas funciones, también dependerá de θ .

No olvidemos que estamos haciendo todo esto para poder estimar θ . Es decir, estamos buscando una expresión para estimar θ . Como L depende de θ , y L es la probabilidad de que ocurra $X_1 = x_1 \wedge X_2 = x_2 \wedge \dots \wedge X_n = x_n$, entonces la probabilidad de que ocurra $X_1 = x_1 \wedge X_2 = x_2 \wedge \dots \wedge X_n = x_n$ depende de θ .

La táctica usada por el método de máxima verosimilitud es, entre todas las expresiones posibles que se pueden proponer para estimar θ , proponer aquella que hace que L sea máxima, y como L es la probabilidad de que ocurra lo que ocurre (es decir, que salgan los determinados valores de la muestra) entonces la expresión para θ que se propone al hacer eso es la que supone que ocurre "lo que más

probabilidades tenía de ocurrir". (De ahí el nombre "máxima verosimilitud"). Para encontrar la expresión para θ que cumpla con dicha condición, es necesario encontrar el θ tal que $L(\theta)$ es máxima. En la mayor parte de los casos, esto se puede hacer derivando L respecto de θ y encontrando el θ tal que la derivada sea cero.

Debido a que L es una productoria, puede resultar muy práctico trabajar con el logaritmo natural de L en vez de con L . Es válido trabajar con $\ln(L)$ en vez de L porque como $\ln(L)$ es monótona y estrictamente creciente con L , entonces donde L tenga su máximo, lo tendrá también $\ln(L)$. Trabajar con $\ln(L)$ es más práctico, porque las propiedades de los logaritmos nos permitirán reemplazar productorias por sumatorias y potencias por factores, y eso por lo general nos simplificará mucho el trabajo a la hora de derivar.

Una vez igualada a cero la derivada, despejamos el parámetro, y esa expresión que nos queda es el estimador de máxima verosimilitud para ese parámetro de esa distribución.

En rigor, es justamente al imponer la condición de que la derivada valga cero que estamos usando la hipótesis de máxima verosimilitud de que ocurre lo que más probabilidades tenía de ocurrir. Por lo tanto, a partir de que se iguala la derivada a cero, lo que queda en la expresión no es θ sino ya su estimador $\hat{\theta}$. Entonces en el ejemplo que dimos más arriba, a partir de donde se dijo "igualamos la derivada a cero" y hasta el final del ejercicio, donde aparece p debería en realidad aparecer \hat{p} . No lo hicimos para no complicar la notación del ejemplo, pero conceptualmente es preferible hacerlo.

Casos patológicos

Puede ocurrir que al derivar L o su logaritmo, e igualar la derivada a cero, no se logre despejar θ en función de las x_i . Eso ocurre cuando la función de verosimilitud no tiene máximos locales, por lo cual la derivada nunca da cero, y al forzar la igualación a cero llegamos a resultados imposibles.

Cuando ocurre eso, por ejemplo en una distribución uniforme, nunca podremos llegar, por medio de la derivación, a una expresión para el parámetro en función de los x_i . Y como justamente construir un estimador es buscar una cuenta que se pueda hacer con los x_i para estimar el parámetro a partir de la muestra, entonces el método tal cual lo hemos descripto no sirve.

Debemos entonces emplear una visión más general del método de máxima verosimilitud:

1) Escribir la función de verosimilitud:

$$L = \prod_{i=1}^n P_X(x_i)$$

si X es discreta

$$L = \prod_{i=1}^n f_X(x_i)$$

si X es continua

- 2) Observar L y obtener la condición sobre el parámetro para que L sea lo más grande posible. Esta condición puede ser, por ejemplo, "que θ se acerque lo más posible a cero", "que θ sea lo más grande posible", etc.
- 3) Revisar la muestra y entre todos los valores posibles de θ que podemos elegir sin que nos quede imposible algún valor de la muestra, elegir el que mejor cumple con la condición determinada en el punto 2.

Ejemplo

Encontrar el estimador de máxima verosimilitud del parámetro b de la siguiente distribución:

$$f_X(x) = \begin{cases} \frac{1}{b} & 0 < x < b \\ 0 & \forall \text{ otro } x \end{cases}$$

Procedemos como en el método que enunciamos en primer lugar:

$$L = \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n \frac{1}{b} = \frac{1}{b^n}$$

- 1)
- 2) $A = \ln(L) = (-n) \ln(b)$
 $\frac{dA}{db} = \frac{-n}{b} = 0$
- 3)
- 4) Vemos que no se puede despejar b en función de los x_i , en este caso porque la función ni siquiera depende de los x_i . De hecho como la distribución es uniforme, ni siquiera la función de densidad dependía de los x_i .

Luego, necesitaremos proceder con el método general:

$$L = \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n \frac{1}{b} = \frac{1}{b^n}$$

- 1)
- 2) Observando $L(b)$, podemos apreciar que crece a medida que b se acerca a cero. Luego, teniendo en cuenta el objetivo de maximizar L, sacaremos la conclusión de que intentaremos que b sea lo más chico posible.
- 3) Para que ningún valor de la muestra nos quede "imposible", todos los valores de la muestra tienen que quedar dentro del dominio de la función de densidad. Y como dicho dominio va entre 0 y b, si b fuera menor que algún elemento de la muestra entonces ese valor de la muestra sería "imposible". Pero no puede ser imposible, porque si pertenece a la muestra entonces de hecho ocurrió. Entonces b tiene que ser necesariamente mayor o igual al mayor elemento de la muestra. Es decir:

$$b \geq \max_{1 \leq i \leq n} \{x_i\}$$

Según la condición hallada en 2), b debe ser lo más chica posible, con lo cual

elegimos $b = \max_{1 \leq i \leq n} \{x_i\}$ y en conclusión el estimador de máxima verosimilitud para el

parámetro b de la distribución estudiada es $\hat{b} = \max_{1 \leq i \leq n} \{x_i\}$

Problemas típicos

1) Encontrar el estimador de máxima verosimilitud para:

- a) el parámetro p de una distribución geométrica, y también su media y su varianza.
- b) el parámetro μ de una distribución de Poisson, y también su media y su varianza.
- c) el parámetro λ de una distribución exponencial negativa, y también su media y su varianza.

Resolución

$$a) P(X = x) = p(1-p)^{x-1}$$

$$L = \prod_{i=1}^n P_X(x_i) = \prod_{i=1}^n p(1-p)^{x_i-1}$$

$$\begin{aligned} A = \ln(L) &= \sum_{i=1}^n \ln(p(1-p)^{x_i-1}) = \sum_{i=1}^n [\ln(p) + (x_i-1)\ln(1-p)] = \sum_{i=1}^n [\ln(p) + x_i \ln(1-p) - \ln(1-p)] = \\ &= \sum_{i=1}^n \ln(p) + \sum_{i=1}^n x_i \ln(1-p) - \sum_{i=1}^n \ln(1-p) = n \ln(p) + \ln(1-p) \sum_{i=1}^n x_i - n \ln(1-p) \end{aligned}$$

$$\begin{aligned} \frac{dA}{dp} &= \frac{n}{p} - \frac{\sum_{i=1}^n x_i}{1-p} + \frac{n}{1-p} = 0 \Rightarrow \frac{n}{\hat{p}} = \frac{\sum_{i=1}^n x_i}{1-\hat{p}} - \frac{n}{1-\hat{p}} \Rightarrow \frac{n}{\hat{p}} = \frac{\left(\sum_{i=1}^n x_i\right) - n}{1-\hat{p}} \Rightarrow \frac{1-\hat{p}}{\hat{p}} = \frac{\left(\sum_{i=1}^n x_i\right) - n}{n} \Rightarrow \\ &\Rightarrow \frac{1}{\hat{p}} - 1 = \frac{\left(\sum_{i=1}^n x_i\right) - n}{n} \Rightarrow \hat{p} = \frac{n}{\left(\sum_{i=1}^n x_i\right)} = \frac{1}{\bar{X}} \end{aligned}$$

Para encontrar el estimador de máxima verosimilitud de la esperanza de la distribución geométrica, podemos usar el estimador que obtuvimos para p , y usar el hecho de que $E(X) = 1/p$. Queda:

$$E(X) = \frac{1}{p} \Rightarrow \hat{E}(X) = \frac{1}{\hat{p}} = \frac{1}{\frac{1}{\bar{X}}} = \bar{X}$$

Análogamente, para la varianza podemos hacer:

$$\sigma_x^2 = \frac{1-p}{p^2} \Rightarrow \sigma_x^2 = \frac{1-\hat{p}}{\hat{p}^2} = \frac{1-\frac{1}{\bar{X}}}{\left(\frac{1}{\bar{X}}\right)^2}$$

Los estimadores para la media y la varianza que se hallan a partir del estimador de máxima verosimilitud para el parámetro, son efectivamente los estimadores de máxima verosimilitud para la media y la varianza.

$$P(X = x) = \frac{e^{-\mu} \mu^x}{x!}$$

b)

$$L = \prod_{i=1}^n P_X(x_i) = \prod_{i=1}^n \frac{e^{-\mu} \mu^{x_i}}{x_i!}$$

$$A = \ln(L) = \sum_{i=1}^n \ln\left(\frac{e^{-\mu} \mu^{x_i}}{x_i!}\right) = \sum_{i=1}^n [\ln(e^{-\mu}) + \ln(\mu^{x_i}) - \ln(x_i!)] = \sum_{i=1}^n [(-\mu) \ln(e) + x_i \ln(\mu) - \ln(x_i!)] =$$

$$= \sum_{i=1}^n [(-\mu) + x_i \ln(\mu) - \ln(x_i!)] = \sum_{i=1}^n [(-\mu)] + \sum_{i=1}^n [x_i \ln(\mu)] - \sum_{i=1}^n [\ln(x_i!)] = -n\mu + \ln(\mu) \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!)$$

$$\frac{dA}{d\mu} = -n + \frac{\sum_{i=1}^n x_i}{\mu} = 0 \Rightarrow \hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} = \bar{X}$$

$$E(X) = \mu \Rightarrow \hat{E}(X) = \hat{\mu} = \bar{X}$$

$$\sigma_x^2 = \mu \Rightarrow \hat{\sigma}_x^2 = \hat{\mu} = \bar{X}$$

c)

$$f_X = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$L = \prod_{i=1}^n f_X(x_i)$$

$$A = \ln(L) = \sum_{i=1}^n \ln(\lambda e^{-\lambda x_i}) = \sum_{i=1}^n [\ln(\lambda) + (-\lambda x_i) \ln(e)] = \sum_{i=1}^n \ln(\lambda) - \sum_{i=1}^n \lambda x_i = n \ln(\lambda) - \lambda \sum_{i=1}^n x_i$$

$$\frac{dA}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0 \Rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{X}}$$

$$E(X) = \frac{1}{\lambda} \Rightarrow \hat{E}(X) = \frac{1}{\hat{\lambda}} = \frac{1}{\frac{1}{\bar{X}}} = \bar{X}$$

$$\sigma_x^2 = \frac{1}{\lambda^2} \Rightarrow \hat{\sigma}_x^2 = \frac{1}{\hat{\lambda}^2} = \frac{1}{\left(\frac{1}{\bar{X}}\right)^2} = \bar{X}^2$$

2) Encontrar el estimador de máxima verosimilitud para el parámetro a de:

$$f_X(x) = \begin{cases} \frac{2(2-x)}{(2-a)^2} & a < x < 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

Resolución

Comenzamos por construir la función de verosimilitud:

$$L(a) = \prod_{i=1}^n \frac{2(2-x_i)}{(2-a)^2}$$

Tomamos el logaritmo y simplificamos:

$$A = \ln(L) = \sum_{i=1}^n \ln\left(\frac{2(2-x_i)}{(2-a)^2}\right) = \sum_{i=1}^n [\ln(2) + \ln(2-x_i) - 2\ln(2-a)] = n \ln(2) + \sum_{i=1}^n \ln(2-x_i) - 2n \ln(2-a)$$

Derivamos con respecto al parámetro a:

$$\frac{dA}{da} = \frac{2n}{2-a}$$

Vemos que esa expresión no depende de las x_i , luego es inútil que sigamos adelante igualando a cero, porque al no depender la función la expresión de las x_i , jamás podremos despejar a en función de las x_i . Nos damos cuenta entonces de que es un caso patológico.

Retrocedamos al objetivo general de maximizar la función de verosimilitud:

$$L(a) = \prod_{i=1}^n \frac{2(2-x_i)}{(2-a)^2}$$

Observamos que cuanto más se acerque a al valor 2, menor será el denominador, y consecuentemente mayor será la función de verosimilitud.

Por otra parte, como la distribución asigna probabilidad no nula a los valores entre a y 2, para que ningún valor de la muestra nos quede imposible, a tiene que ser menor que el mínimo valor de la muestra. Es decir, $a \leq \min\{x_i\}$.

Y como habíamos establecido, para maximizar la función de verosimilitud en este caso necesitamos que a esté lo más cerca del 2 posible. En consecuencia, nos quedamos con el valor $a = \min\{x_i\}$.

3) Encontrar el estimador de máxima verosimilitud para el parámetro λ de:

$$f_x(x) = \begin{cases} \frac{1}{2\sqrt{\lambda x}} & 0 < x < \lambda \\ 0 & \forall \text{ otro } x \end{cases}$$

Resolución

Comenzamos por construir la función de verosimilitud:

$$L(\lambda) = \prod_{i=1}^n \frac{1}{2\sqrt{\lambda x_i}}$$

Tomamos el logaritmo y simplificamos:

$$\begin{aligned} A = \ln(L) &= \sum_{i=1}^n \ln\left(\frac{1}{2\sqrt{\lambda x_i}}\right) = \sum_{i=1}^n [\ln(2) - \ln(\sqrt{\lambda x_i})] = -\sum_{i=1}^n \left[\ln(2) + \frac{1}{2}(\ln(\lambda) + \ln(x_i))\right] = \\ &= -n \ln(2) - \frac{n \ln(\lambda)}{2} - \frac{1}{2} \sum_{i=1}^n \ln(x_i) \end{aligned}$$

Derivamos con respecto al parámetro λ :

$$\frac{dA}{d\lambda} = \frac{n}{2\lambda}$$

Vemos que es un caso patológico, y retrocedamos al objetivo general de maximizar la función de verosimilitud:

$$L(\lambda) = \prod_{i=1}^n \frac{1}{2\sqrt{\lambda x_i}}$$

Observamos que cuanto más se acerque λ al valor cero, menor será el denominador, y consecuentemente mayor será la función de verosimilitud. Por otra parte, como la distribución asigna probabilidad no nula a los valores entre 0 y λ , para que ningún valor de la muestra nos quede imposible, λ tiene que ser mayor que el máximo valor de la muestra. Es decir, $\lambda \geq \max\{x_i\}$. Y como habíamos establecido, para maximizar la función de verosimilitud en este caso necesitamos que λ esté lo más cerca del cero posible. En consecuencia, nos quedamos con el valor $\lambda = \max\{x_i\}$.

Estimadores más comunes

Presentamos a continuación los estimadores más comunes de determinados parámetros.

Estimador para la media de una población

Para estimar la media de una población, se utiliza el promedio muestral, cuya fórmula es:

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

Sus propiedades ya fueron analizadas en parte en la primera sección de este capítulo, se determinó que es insesgado y eficiente.

Estimador para la varianza y el desvío de una población

Para estimar la varianza poblacional se utiliza la varianza muestral, que se nota S^2 y cuya fórmula es:

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

El estimador varianza muestral es insesgado y eficiente.

Para estimar el desvío poblacional se usa el desvío muestral, que es la raíz cuadrada de la varianza muestral. Su fórmula es:

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x - x_i)^2}{n - 1}}$$

Estimador para la proporción de una muestra de tamaño grande

Debemos aquí entender proporción de una forma parecida a probabilidad. La proporción es por ejemplo lo que usamos como parámetro p de una binomial cuando nos dicen "el 80% de la gente tiene ojos oscuros". Su estimador se nota \hat{p} y su fórmula es:

$$\hat{p} = \frac{X}{n}$$

donde n es el tamaño de la muestra y X es la cantidad de éxitos obtenidos.

Problemas típicos

1) Se toman 10 piezas fabricadas por una máquina, para estudiar el peso de las piezas producidas por dicha máquina. Los pesos de la muestra de 10 piezas son:

156, 143, 185, 182, 202, 191, 152, 168, 117, 199.

Estime la media, la varianza y el desvío de la distribución del peso de las piezas que produce la máquina.

Resolución:

El promedio muestral es un buen estimador de la media y como no tenemos más información sobre la población, lo usaremos.

$$E(X) \approx \bar{X} = \frac{\sum_{i=1}^n x_i}{n} = 170.2$$

También usaremos la varianza y el desvío muestral para estimar la varianza y el desvío poblacional:

$$\sigma_x^2 \approx S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1} = 726$$

$$\sigma_x = \sqrt{\sigma_x^2} \approx \sqrt{S_x^2} = 27$$

2) En un control de calidad, se verificaron 1000 piezas, de las cuales 9 resultaron defectuosas. Indique la probabilidad de que en un lote de 6 piezas haya alguna defectuosa.

Resolución

Este ejercicio resulta parecido a los que resolvimos cuando estudiamos la distribución binomial. La diferencia es que en ese capítulo el dato era: "el 0,9% de las piezas son defectuosas", o "9 de cada 1000 piezas son defectuosas". Ahora estamos bajando ese problema a la vida real, haciendo énfasis en que el 0,9% que nos daban de dato no es en rigor cierto, sino que se calculó a partir de una muestra.

Cuando la muestra es grande, \hat{p} es un muy buen estimador de p , por eso la p que

$$\hat{p} = \frac{X}{n} = \frac{9}{1000} = 0,009$$

usaremos con la binomial será en realidad

Luego si X es la cantidad de piezas defectuosas en un lote de 6, la probabilidad de que haya piezas defectuosas será $1 - P(X = 0)$, donde $X:Bi(n = 6 ; p = 0,009)$.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 12 de julio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 22 de julio de 2004

CAPÍTULO IX

Intervalos de confianza

En el capítulo anterior estudiamos los estimadores. Vimos que los estimadores son una determinada función de los resultados que se obtienen en un experimento, y que nos sirven para estimar el valor de algún determinado parámetro. Estudiamos asimismo cómo evaluarlos para saber si eran "buenos" o "malos".

Sin embargo, el hecho de saber que un estimador es "bueno" o "malo" no nos da una idea numérica de cuánto es el error que podemos estar cometiendo al usar el estimador para estimar el parámetro. Dicho de otra forma, no nos da una idea concreta de cuánta (en términos numéricos) confianza le podemos tener.

Para solucionar este problema, hará su aparición en escena otra herramienta: el intervalo de confianza. El mismo es un intervalo entre dos números, tal que hay una determinada probabilidad de que el verdadero valor del parámetro (desconocido) pertenezca al intervalo.

En rigor, deberíamos decir que el intervalo tiene una determinada probabilidad de contener al verdadero valor del parámetro. Esto es así porque en realidad el valor del parámetro, aunque es desconocido, existe y es una constante. Lo que son variables aleatorias son los límites del intervalo. Por eso en rigor es más correcto decir que es el intervalo el que tiene una determinada probabilidad de contener al verdadero valor, y no que es el verdadero valor el que tiene una determinada probabilidad de estar contenido en el intervalo.

De todos modos, más allá de la discusión filosófica, la probabilidad de la que estamos hablando se denomina nivel de confianza (NC).

Por ejemplo, dada una población normal cuyo parámetro μ desconocemos, nos pueden pedir que indiquemos el intervalo de confianza del 95% para estimarla. Esto significa que debemos dar dos números L_1 y L_2 tales que haya un 95% de probabilidad de que el verdadero valor de μ se encuentre entre ellos. Es decir, que haya un 95% de probabilidad de que el intervalo que determinemos contenga al verdadero valor de μ . Dicho de otro modo, debemos encontrar los límites tal que haya un nivel del 95% de confianza de que el intervalo que determinemos contenga al verdadero valor de μ .

Cuanto mayor sea el nivel de confianza que se pida, mayor será la longitud del intervalo que se construya, para poder garantizar la probabilidad pedida de que el verdadero valor esté contenido dentro del intervalo dado.

Cuanto mayor sea el tamaño de la muestra, menor será la longitud del intervalo de confianza, porque los estimadores que usemos se parecerán más a los parámetros que desconocemos.

En general, nos interesa obtener intervalos lo más pequeños posibles, para tener una idea más precisa del valor del parámetro que desconocemos.

Intervalo de confianza para la media de una población

Población	Desvío poblacional conocido (usamos σ)	Desvío poblacional desconocido (usamos S , el desvío muestral)
normal	$L_{1;2} = \bar{X} \pm \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$	$L_{1;2} = \bar{X} \pm \frac{t_{1-\frac{\alpha}{2}; n-1} S}{\sqrt{n}}$
no normal	$L_{1;2} = \bar{X} \pm \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$ debemos pedir $n \geq 30$	$L_{1;2} = \bar{X} \pm \frac{z_{1-\frac{\alpha}{2}} S}{\sqrt{n}}$ debemos pedir $n \geq 50$

donde:

- \bar{X} es el promedio muestral.
 - n es el tamaño de la muestra.
 - σ es el desvío poblacional.
 - S es el desvío muestral (calculado a partir de la muestra).
 - $\alpha = 1 - \text{NC}$. Es decir, si nos piden 95% de confianza, $\alpha = 1 - 0,95 = 0,05$.
 - $z_{1-\frac{\alpha}{2}}$ es un fractil de la distribución normal estándar tal como lo estudiamos en el capítulo 6.
 - $t_{1-\frac{\alpha}{2}; n-1}$ es un fractil de la distribución t-Student tal como lo estudiamos en el capítulo 7.
- Los fractiles z y t se obtienen de las correspondientes tablas.

Explicación y justificación

Comenzaremos por estudiar el caso en que la población es normal y conocemos el desvío poblacional (σ), y queremos estudiar mediante una muestra el valor del parámetro μ de la distribución.

Según la definición que dimos de intervalo de confianza, construir dicho intervalo consiste en proveer los límites L_1 y L_2 tales que la probabilidad de que contengan al verdadero valor del parámetro (es decir, lo que **no** sabemos) sea una determinada probabilidad alta que nos piden, por ejemplo, 90%. A esa probabilidad alta con la que debemos cumplir, en este caso 90%, la denominamos nivel de confianza (NC).

Para hacer las cuentas, vamos a trabajar con $\alpha = 1 - NC$.

Es decir, estamos buscando L_1 y L_2 tales que:

$$P(L_1 < \mu < L_2) = NC = 1 - \alpha$$

Vimos que para estimar μ se usa \bar{X} . ¿Cómo está distribuido \bar{X} ? Es el promedio de los n elementos extraídos de una población normal. Como vimos en el capítulo anterior, el valor de cada uno de esos resultados tomados de una población es una variable aleatoria normal, cuyos media y desvío son los mismos que los de la población de la cual vienen, es decir, la media de cada x_i es el μ desconocido, y el desvío de cada x_i es el σ que tenemos como dato. Por lo tanto \bar{X} es un promedio de variables aleatorias normales. Y el promedio es un caso particular de combinación lineal (donde todos los coeficientes valen $1/n$). Como la combinación lineal de variables normales es normal, \bar{X} también tiene una distribución normal. Determinemos sus parámetros:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n}x_1 + \frac{1}{n}x_2 + \dots + \frac{1}{n}x_n$$

$$E(\bar{X}) = E\left(\frac{1}{n}x_1 + \frac{1}{n}x_2 + \dots + \frac{1}{n}x_n\right) = \frac{1}{n}E(x_1) + \frac{1}{n}E(x_2) + \dots + \frac{1}{n}E(x_n) = \frac{n\mu}{n} = \mu$$

$$\sigma_{\bar{X}}^2 = \sqrt{\sum_{i=1}^n \frac{1}{n^2} \sigma_x^2} = \frac{\sigma}{\sqrt{n}}$$

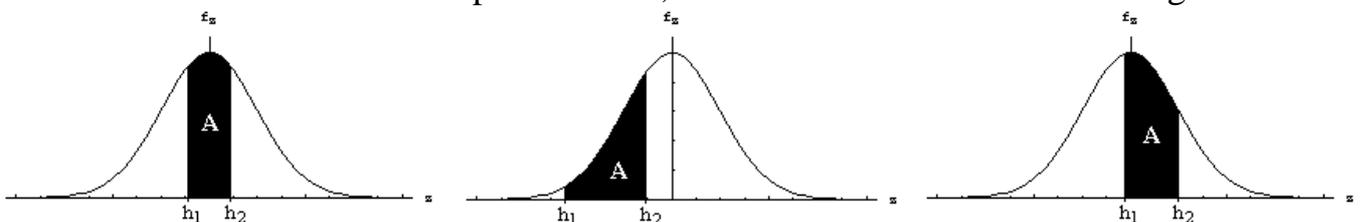
$$\bar{X} : N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

Con lo cual

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Por lo tanto, como se vio en el capítulo 6, si tomamos entonces Z tiene una distribución normal estándar.

Si quisiéramos encontrar dos valores h_1 y h_2 tales que la probabilidad de que Z esté entre ellos sea una determinada probabilidad, tendríamos infinitas formas de lograrlo:



En los gráficos vemos 3 de las infinitas formas posibles de tomar h_1 y h_2 tales que la probabilidad de que Z esté entre ellos (el área encerrada) valga A .

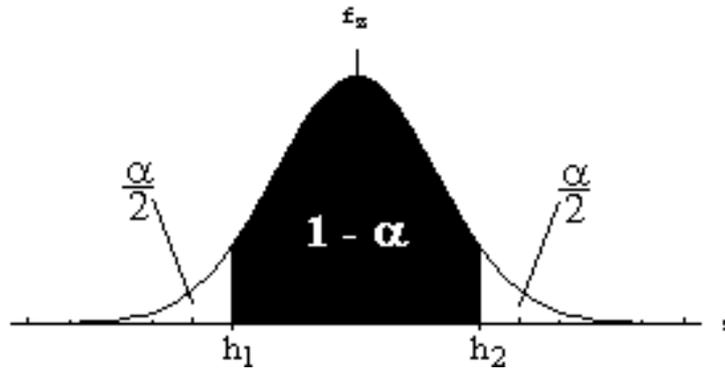
De todas esas infinitas formas de elegir h_1 y h_2 , nos quedaremos con la representada por el gráfico de la izquierda, es decir, aquella en la cual el área es simétrica respecto del eje vertical. Esto lo hacemos por dos razones:

- porque es la que opción que nos proporciona h_1 y h_2 más cercanos (intervalo más chico).
- porque es lo más natural (lo entenderemos intuitivamente).

Entonces, decidiendo que el área será simétrica alrededor del cero (es decir, $h_1 = -h_2$)

y especificando el área que queremos que quede encerrada (NC), podemos encontrar h_1 y h_2 .

Si queremos que el área encerrada sea NC, es decir, $1-\alpha$, nos queda:



El área acumulada a la izquierda de h_2 es la suma de $1-\alpha$ (el área encerrada) y $\alpha/2$ (el área a la izquierda de h_1). Entonces el área acumulada a la izquierda de h_2 es:

$$1 - \alpha + \alpha/2 = 1 - \alpha/2$$

Por lo tanto h_2 no es otra cosa que el fractil $Z_{1-\alpha/2}$, y $h_1 = -Z_{1-\alpha/2}$ porque $h_1 = -h_2$.
Entonces escribimos:

$$P\left(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}\right) = 1 - \alpha$$

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

Y como dijimos que entonces podemos escribir:

$$P\left(-Z_{1-\alpha/2} < \frac{X - \mu}{\frac{\sigma}{\sqrt{n}}} < Z_{1-\alpha/2}\right) = 1 - \alpha$$

Despejando μ del centro de la doble inecuación, queda:

$$P\left(X - \frac{Z_{1-\alpha/2} \sigma}{\sqrt{n}} < \mu < X + \frac{Z_{1-\alpha/2} \sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$\begin{array}{ccc} \bar{X} - \frac{Z_{1-\alpha/2} \sigma}{\sqrt{n}} & & \bar{X} + \frac{Z_{1-\alpha/2} \sigma}{\sqrt{n}} \\ \hline L_1 & \bar{X} & L_2 \end{array}$$

Con lo cual hemos encontrado L_1 y L_2 tales que, dada la muestra, y dado α , la probabilidad de que μ se encuentre entre L_1 y L_2 vale $1-\alpha$. Y eso es lo que buscábamos. De ahí obtenemos:

$$L_{1;2} = \bar{X} \pm \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$$

Faltan estudiar los casos de que la población no sea normal, o no conozcamos el desvío o las dos cosas.

Si la población no es normal, pero n es grande, no hay ningún problema, porque el teorema central del límite nos garantiza que \bar{X} , que en la deducción anterior supusimos normal, tiene una distribución aproximadamente normal. Entonces para $n \geq 30$ podemos usar para encontrar el intervalo de confianza de la media de una población no normal la misma fórmula que usamos para una población normal.

Si no conocemos el desvío poblacional (σ), nos veremos obligados a usar el desvío muestral (S). Dependiendo del tamaño de la muestra, S será una mejor o peor aproximación de σ , pero de hecho no es σ , porque σ es una constante y S es una variable aleatoria (porque es una función de variables aleatorias). Como $\sigma \neq S$,

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

entonces tomando $\frac{S}{\sqrt{n}}$, Z **no** es una variable normal estándar. Es algo **parecido**, pero en rigor no lo es. Su distribución recibe el nombre de "t de student" o bien "t-student", y fue estudiada en el capítulo 7. Dicha distribución, a diferencia de la normal estándar, depende de n , por eso para calcular su fractil necesitamos no solamente el área acumulada sino también el valor de n .

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

Por lo tanto, si tomamos $\frac{S}{\sqrt{n}}$, donde n es el tamaño de una muestra cuya media es μ , \bar{X} es el promedio muestral y S el desvío muestral, entonces T tiene una distribución t-Student con parámetro $\nu = n - 1$. La distribución t-Student también es simétrica, con lo cual podemos usar, al igual que hicimos antes con la normal, que h_1

$$= -h_2, \text{ con lo cual } t_{\frac{\alpha}{2}; n-1} = -t_{1-\frac{\alpha}{2}; n-1}.$$

El fractil $t_{1-\frac{\alpha}{2}; n-1}$ puede obtenerse de su correspondiente tabla, al igual que los fractiles de la normal estándar.

En resumen, al no conocer el desvío poblacional, cambiamos σ por S , lo cual

también nos obliga a cambiar $z_{1-\frac{\alpha}{2}}$ por $t_{1-\frac{\alpha}{2}; n-1}$, y queda:

$$L_{1;2} = \bar{X} \pm \frac{t_{1-\frac{\alpha}{2}; n-1} S}{\sqrt{n}}$$

¿Qué pasaría si debiéramos ocuparnos de un caso en el cual ocurriera simultáneamente que la población no fuera normal y que no conociéramos el desvío poblacional?

Vimos que si la población no es normal, alcanza con pedir n grande para que se pueda usar la misma fórmula. Y que si no conocemos el desvío, usamos S y

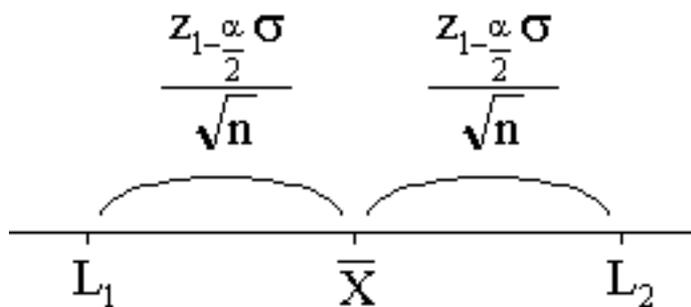
$t_{1-\frac{\alpha}{2}; n-1}$ en vez de σ y $z_{1-\frac{\alpha}{2}}$. Pero en el capítulo 7 vimos que cuando n es lo suficientemente grande, la distribución t-Student se parece mucho a la distribución

normal, con lo cual $t_{1-\frac{\alpha}{2}; n-1} \approx z_{1-\frac{\alpha}{2}}$.

Luego, si $n \geq 50$, podemos usar:

$$L_{1;2} = \bar{X} \pm \frac{z_{1-\frac{\alpha}{2}} S}{\sqrt{n}}$$

Errores



Si estamos estimando μ con \bar{X} , y μ resulta estar dentro del intervalo de confianza que construimos, entonces como \bar{X} está en el centro del intervalo, el máximo error que puede tener \bar{X} con respecto a μ se da en el caso en que μ coincida con uno de los límites del intervalo.

Entonces, por ejemplo, en el caso de que la población sea normal y σ sea conocido, y por lo tanto siendo la fórmula que utilizamos:

$$L_{1;2} = \bar{X} \pm \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$$

entonces el máximo error que podemos cometer es $\frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$.

Es decir, el máximo error posible coincide con el término al que le aplicamos el \pm .

Esto se cumple también para los otros 3 casos que hemos estudiado.

Longitud del intervalo

Mirando las fórmulas halladas, observamos que efectivamente, como dijimos al comienzo del capítulo, cuanto más grande es el tamaño de la muestra(n) más pequeño queda el intervalo, y cuanto más confianza queremos tener, más grande es el intervalo.

Problemas típicos

1) La duración de unas determinadas baterías es una variable aleatoria normal, cuya media se desea estimar, para lo cual se toma una muestra de 9 baterías, cuyas duraciones, en horas, resultan: 6.3, 6.8, 7.3, 5.4, 8.1, 7.9, 6.9, 6.2, 8.3.

Se pide:

- a) Calcular el intervalo del 95% de confianza para estimar la media.**
- b) Suponga que se averigua que el desvío poblacional de la duración es 1, ¿cómo sería el intervalo de confianza?**
- c) Si se desea reducir un 20% el intervalo obtenido en b), manteniendo el nivel de confianza, ¿cuántas baterías adicionales se deberían probar?**
- d) Si se desea que el máximo error sea 0.15, ¿cuántas baterías se deberían probar en total?**
- e) Volviendo a la situación que teníamos en b), ¿cuántas baterías adicionales debemos revisar, para aumentar la confianza al 99%?**

Resolución

a) Como no conocemos el desvío poblacional, usaremos

$$L_{1;2} = \bar{X} \pm \frac{t_{1-\frac{\alpha}{2}; n-1} S}{\sqrt{n}}$$

El tamaño de la muestra es $n = 9$.

Vamos a necesitar calcular \bar{X} y S .

$$\bar{X} = 7.02$$

$$S_x = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = 0.97$$

$$\alpha = 1 - NC = 0.05$$

$$1 - \frac{\alpha}{2} = 0.975$$

Buscamos el fractil de la t-Student en la tabla, y obtenemos $t_{0.975;8} = 2.306$.

Ya estamos en condiciones de obtener los límites del intervalo:

$$L_{1;2} = 7.02 \pm \frac{2.306 \cdot 0.97}{\sqrt{9}}$$

El intervalo de confianza nos queda: (6.28 ; 7.77).

b) Como conocemos el desvío poblacional, ahora vamos a usar:

$$L_{1;2} = \bar{X} \pm \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$$

Buscamos el fractil de la normal estándar en la tabla, y obtenemos $z_{0.975} = 1.96$
 El intervalo de confianza nos queda: (6.37 ; 7.68).

Vemos que aunque el desvío poblacional es un poco mayor que el desvío muestral que habíamos utilizado, de todas formas el intervalo nos quedó más chico. Esto es porque si conocemos el desvío poblacional, tenemos más información sobre el parámetro desconocido, y entonces podemos garantizar la misma probabilidad, con un intervalo más chico.

c) La longitud del intervalo obtenido en b es $2 \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$. Nos piden que la longitud del nuevo intervalo (aumentando n) sea el 80% del actual. Planteamos:

$$2 \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} \cdot 0.8 = 2 \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n'}} \quad (\text{la nueva longitud es 0.8 por la longitud anterior})$$

Es evidente que en el caso general, el nuevo valor de n no va a ser entero. Por eso en realidad tenemos que interpretar que nos piden que el intervalo se reduzca "al menos" un 20%. Es decir, si llegamos a un valor no entero de n, debemos redondear para arriba para que nuestra respuesta cumpla con lo que piden. O para evitarnos el problema de tener que pensar para qué lado redondear, podemos plantear directamente la inecuación que refleja la interpretación:

$$2 \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} \cdot 0.8 \geq 2 \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n'}}$$

despejamos n' y queda:

$$\sqrt{n'} \geq \frac{\sqrt{n}}{0.8} = 3.75 \Rightarrow n' \geq 14.06 \Rightarrow n' = 15$$

Vemos que efectivamente redondeamos para arriba como habíamos pensado en un principio.

Si se necesitan revisar 15 baterías para cumplir con lo que piden, como nos preguntaron cuántas adicionales a las que ya se habían revisado se deben revisar, nuestra respuesta debe ser 6.

$$\frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$$

d) El máximo error es, como habíamos analizado, $\frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$. Nos piden que el máximo error sea 0.15. Nuevamente debemos interpretar: como el n no nos va a quedar entero, debemos redondear para arriba, porque a mayor n, menor error, por lo tanto si redondeamos para abajo el máximo error nos va a quedar mayor que 0.15. O podemos directamente plantear la inecuación correspondiente:

$$\frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n'}} \leq 0.15$$

(porque debemos interpretar "como máximo" 0.15)

Despejando n', queda:

$$\sqrt{n'} \geq \frac{z_{1-\frac{\alpha}{2}} \sigma}{0.15} = 13.067 \Rightarrow n' \geq 170.74 \Rightarrow n' = 171$$

Vemos que redondeamos para arriba, como habíamos pensado.

También observamos que n es bastante más grande que los valores de n que veníamos manejando, con lo cual vemos que obtener intervalos pequeños con altos niveles de confianza requiere tomar muestras de tamaños grandes.

e) En b) obtuvimos que con $n = 9$, se obtuvo, para 95% de confianza, que

$$\frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} = 0.65$$

, con lo cual el intervalo nos quedó (6.37 ; 7.68).

Ahora queremos que ese mismo intervalo tenga el 99% de confianza, para lo cual aumentaremos n. El nuevo α vale 0.01. Para que se cumpla lo que piden, debemos escribir la siguiente inecuación:

$$\frac{z_{1-\frac{0.01}{2}} \sigma}{\sqrt{n'}} \leq 0.65$$

(porque debemos interpretar "al menos" el 99% de confianza).

Buscando el fractil y despejando n', obtenemos:

$$\sqrt{n'} \geq \frac{z_{1-\frac{0.01}{2}} \sigma}{0.65} = 3.96 \Rightarrow n' \geq 15.7 \Rightarrow n' = 16$$

Intervalo de confianza para la varianza y el desvío de una población

Intervalo de confianza de la varianza de una población

$$L_1 = \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2};n-1}} \quad L_2 = \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2};n-1}}$$

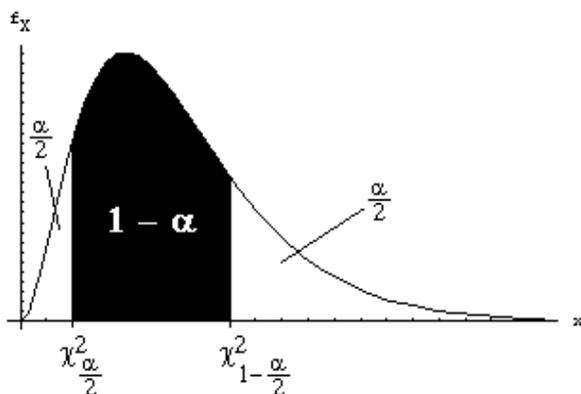
donde:

- n es el tamaño de la muestra. Si la población no es normal, debe pedirse $n \geq 30$.
- S^2 es la varianza muestral.
- $\alpha = 1 - \text{NC}$. Es decir, si nos piden 95% de confianza, $\alpha = 1 - 0,95 = 0,05$.
- $\chi^2_{1-\frac{\alpha}{2};n-1}$, $\chi^2_{\frac{\alpha}{2};n-1}$ son fractiles de la distribución chi-cuadrado con $n-1$ grados de libertad, que se obtienen de la tabla.

Intervalo de confianza para el desvío de una población

Los límites L_1 y L_2 son las raíces cuadradas de los límites dados para el intervalo de la varianza.

Explicación y justificación



$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

Si tomamos $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$, donde n es el tamaño de una muestra tomada de una población normal, S^2 es la varianza muestral y σ^2 es la varianza poblacional,

entonces χ^2 tiene una distribución chi-cuadrada con grados de libertad $v = n - 1$. La distribución chi-cuadrada fue estudiada en el capítulo 7.

Luego análogamente a como hicimos en la sección anterior, queremos obtener L_1 y L_2 tales que:

$$P(L_1 < \sigma^2 < L_2) = 1 - \alpha$$

Similarmente a como hicimos con la Z , escribimos que:

$$P\left(\chi_{\frac{\alpha}{2};v}^2 < X^2 < \chi_{1-\frac{\alpha}{2};v}^2\right) = 1 - \alpha$$

(la X^2 no es la letra "equis" sino "chi" mayúscula)

Luego reemplazando por el cambio de variables que habíamos tomado, queda:

$$P\left(\chi_{\frac{\alpha}{2};v}^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{1-\frac{\alpha}{2};v}^2\right) = 1 - \alpha$$

Despejando σ^2 del centro de la inecuación, queda:

$$P\left(\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2};v}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2};v}^2}\right) = 1 - \alpha$$

y así llegamos a la fórmula que habíamos dado.

Si la población no es normal, para usar esta fórmula debe pedirse que $n \geq 30$.

Problemas típicos

1) La duración de unas determinadas baterías es una variable aleatoria normal, y de desea estimar su varianza y su desvío. Para eso se toma una muestra de 9 baterías, cuyas duraciones, en horas, resultan: 6.3, 6.8, 7.3, 5.4, 8.1, 7.9, 6.9, 6.2, 8.3.

¿Cuál es el intervalo del 90% de confianza para estimar la varianza y el desvío?

Resolución

Usaremos:

$$L_1 = \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2};n-1}^2} \quad L_2 = \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2};n-1}^2}$$

Calculamos:

$$S_x^2 = \frac{\sum_{i=1}^n (x - x_i)^2}{n-1} = 0.94$$

$$\frac{\alpha}{2} = 0.05 \quad 1 - \frac{\alpha}{2} = 0.95$$

$$\alpha = 1 - NC = 0.1$$

Buscamos los fractiles en la tabla:

$$\chi^2_{0.95;8} = 15.507 \quad \chi^2_{0.05;8} = 2.733$$

Obtenemos el intervalo del 90% de confianza para la varianza

$$L_1 = 0.48 \quad L_2 = 2.75$$

El del desvío se obtiene simplemente tomando las raíces cuadradas de los límites:

$$L_1 = 0.70 \quad L_2 = 1.66$$

También deben considerarse problemas típicos las preguntas hechas en la sección anterior sobre los intervalos para la media. No incluimos nuevamente problemas de ese tipo porque el método es exactamente el mismo.

Intervalo de confianza para una proporción

Se desea conocer una determinada proporción p en una población. Para eso se toma una muestra de tamaño n , contabilizándose X éxitos en esa muestra. Se estimará p mediante $\hat{p} = X / n$.

Se requiere que n sea lo suficientemente grande para que se cumplan:
 $n\hat{p} \geq 5$ y $n(1 - \hat{p}) \geq 5$

Los límites L_1 y L_2 del intervalo de confianza son:

$$L_{1;2} = \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

donde:

- $\hat{p} = \frac{X}{n}$

- n es el tamaño de la muestra.

- X la cantidad de éxitos en esa muestra.

- $\alpha = 1 - \text{NC}$. Es decir, si nos piden 95% de confianza, $\alpha = 1 - 0,95 = 0,05$.

- $z_{1-\frac{\alpha}{2}}$ es un fractil de la normal estándar tal como lo estudiamos en el capítulo 6.

También hay otra fórmula, que es ligeramente más exacta, pero que requiere hacer muchas más cuentas:

$$L_{1;2} = \frac{2n\hat{p} + z_{1-\frac{\alpha}{2}}^2 \pm \sqrt{4nz_{1-\frac{\alpha}{2}}^2(\hat{p} - \hat{p}^2) + z_{1-\frac{\alpha}{2}}^4}}{2(n + z_{1-\frac{\alpha}{2}}^2)}$$

En general no se recomienda su uso, ya que la precisión que se gana no resulta significativamente superior a la de la fórmula dada en primer lugar.

Explicación y justificación

Estamos buscando L_1 y L_2 tales que:

$$P(L_1 < p < L_2) = 1 - \alpha$$

Tomamos una muestra de tamaño n , y obtenemos X éxitos. Para estimar p usamos

$\hat{p} = \frac{X}{n}$, donde X es la cantidad de éxitos en la muestra y n el tamaño de la muestra.

¿Qué es tomar una muestra en este caso? Hacer el experimento de Bernoulli n veces, tras lo cual se obtienen X éxitos. En cada una de las veces que se hizo el experimento la probabilidad de éxito era p (desconocida). Entonces X es una variable aleatoria binomial, con parámetros n (tamaño de la muestra) y p (probabilidad de éxito en cada intento). Tal como se vio en la sección 3 del capítulo 6, si X es una variable aleatoria binomial con parámetros n y p , y

$np \geq 5$ y $n(1-p) \geq 5$ entonces su distribución se puede aproximar por una normal:
 $X: N(n, p; \sqrt{n \cdot p \cdot (1-p)})$

Si $\hat{p} = \frac{X}{n}$ entonces \hat{p} es una función lineal de una variable normal, y consecuentemente también tiene una distribución aproximadamente normal:

$$\hat{p}: N\left(p; \sqrt{\frac{p \cdot (1-p)}{n}}\right)$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot (1-p)}{n}}}$$

Por lo tanto, si tomamos Z , Z tiene una distribución aproximadamente normal estándar.

Luego procedemos igual que en la sección sobre el intervalo de confianza de la media. Escribimos:

$$P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot (1-p)}{n}}}$$

Y como dijimos que entonces podemos escribir:

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\hat{p} - p}{\sqrt{\frac{p \cdot (1-p)}{n}}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Despejando p del centro de la doble inecuación, queda:

$$P\left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p \cdot (1-p)}{n}} < p < \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p \cdot (1-p)}{n}}\right) = 1 - \alpha$$

Con lo cual ya casi encontramos los L_1 y L_2 que buscábamos. Falta ocuparnos de una cuestión: como no conocemos p , no podemos conocer en forma exacta el

desvío poblacional $\sqrt{p \cdot (1-p)}$, que aparece a ambos lados en la doble inecuación. Lo

podemos aproximar mediante $\sqrt{\hat{p} \cdot (1-\hat{p})}$, pero entonces ya no es cierta la estandarización que usamos para conseguir Z , porque ésta requiere el desvío

poblacional. Sin embargo, cuando n es grande, el error que se comete al usar esa aproximación es pequeño, porque \hat{p} es un estimador eficiente y porque la raíz amortigua los errores.

En conclusión, cuando n es grande, el error que queda en el coeficiente que

acompaña al fractil de la normal por haber usado $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ en vez de $\sqrt{\frac{p(1-p)}{n}}$ es pequeño.

Adoptamos entonces dicha aproximación, y queda:

$$P\left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$$

Con lo cual hemos encontrado L_1 y L_2 tales que, dada la muestra, y dado α , la probabilidad de que p se encuentre entre L_1 y L_2 vale $1 - \alpha$. Y eso es lo que buscábamos. De ahí obtenemos:

$$L_{1;2} = \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Y esa es la primera fórmula que dimos. La segunda es un poco más exacta, y se

obtiene evitando la aproximación de $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ por $\sqrt{\frac{p(1-p)}{n}}$. Volvamos a la doble inequación que teníamos antes de usar la aproximación:

$$P\left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

El problema con esa definición del intervalo es que nos queda en función del p desconocido. Pero abstrayéndonos de eso, el límite superior del intervalo es el p tal

$$\text{que: } p = \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \quad \text{y el inferior es el } p \text{ tal que } p = \hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}.$$

Si tomamos cualquiera de las dos expresiones, pasamos \hat{p} restando, y elevamos ambos miembros de la ecuación al cuadrado, para cualquiera de las dos expresiones obtendríamos:

$$(p - \hat{p})^2 = z_{1-\frac{\alpha}{2}}^2 \frac{p(1-p)}{n}$$

Haciendo el cuadrado del binomio, y pasando algunos términos, podemos llegar a una ecuación cuadrática en función de p :

$$p^2 \left(1 + \frac{z_{1-\frac{\alpha}{2}}^2}{n}\right) - \left(2\hat{p} + \frac{z_{1-\frac{\alpha}{2}}^2}{n}\right)p + \hat{p}^2 = 0$$

De donde usando la fórmula para resolver ecuaciones cuadráticas y haciendo algunas simplificaciones obtenemos:

$$L_{1;2} = \frac{2n\hat{p} + z_{1-\frac{\alpha}{2}}^2 \pm \sqrt{4nz_{1-\frac{\alpha}{2}}^2(\hat{p}-\hat{p}^2) + z_{1-\frac{\alpha}{2}}^4}}{2(n+z_{1-\frac{\alpha}{2}}^2)}$$

Problemas típicos

1) De la producción de una fábrica de chips, se revisan 200 chips, encontrándose 12 defectuosos. Halle el intervalo del 99% de confianza para la verdadera proporción de chips defectuosos en la producción de la fábrica.

Resolución

Usaremos:

$$L_{1;2} = \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\hat{p} = \frac{X}{n} = \frac{12}{200} = 0.06$$

Buscamos el fractil de la normal estándar en la tabla, y obtenemos $z_{0.995} = 2.58$.

Reemplazamos los datos en la fórmula:

$$L_{1;2} = 0.06 \pm 2.58 \sqrt{\frac{0.06(1-0.06)}{200}} = 0.06 \pm 2.58 \cdot 0.01679$$

Obtenemos el intervalo de confianza: $(L_1 ; L_2)$, es decir, $(0,017 ; 0,103)$.

También deben considerarse problemas típicos las preguntas hechas en la sección sobre los intervalos para la media. No incluimos nuevamente problemas de ese tipo porque el método es exactamente el mismo.

Intervalo de confianza para la diferencia de dos medias

Se desea el intervalo de confianza para la diferencia $\mu_1 - \mu_2$, donde μ_1 y μ_2 son las medias de dos poblaciones independientes 1 y 2. La fórmula que usaremos depende de la información que tengamos acerca de las varianzas de las poblaciones en cuestión.

Se distinguen 3 casos:

CASO 1: Se conocen las varianzas poblacionales σ_1^2 y σ_2^2 .

CASO 2: No se conocen las varianzas poblacionales.

CASO 3: No se conocen las varianzas poblacionales, pero aunque no se las conoce se sabe que son iguales. Es decir, σ_1^2 y σ_2^2 son desconocidas pero por alguna razón se sabe que $\sigma_1^2 = \sigma_2^2$ (el caso 3 es un caso particular del caso 2, que lleva a fórmulas más simples).

Tanto en el caso 2 como en el caso 3, al no conocerse las varianzas poblacionales se las estima con S_1^2 y S_2^2 . Cabe destacar que en el caso 3, el hecho de que para estimar σ_1^2 y σ_2^2 estemos usando S_1^2 y S_2^2 (que arrojarán valores distintos) no constituye una contradicción, porque como S_1^2 y S_2^2 son calculadas a partir de muestras, debido a la naturaleza aleatoria de las muestras tendrán valores ligeramente distintos aunque las muestras se hayan tomado de poblaciones con igual varianza.

A continuación, las fórmulas:

Varianzas	Fórmula
1) Se conocen las varianzas poblacionales σ_1^2 y σ_2^2	$L_{1;2} = \bar{X}_1 - \bar{X}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
2) No se conocen σ_1^2 y σ_2^2 . Se las estima con las varianzas muestrales S_1^2 y S_2^2 .	$L_{1;2} = \bar{X}_1 - \bar{X}_2 \pm t_{1-\frac{\alpha}{2};v} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \quad \text{con} \quad v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$
3) No se conocen σ_1^2 y σ_2^2 , pero se sabe que son iguales. Se las estima con S_1^2 y S_2^2 .	$L_{1;2} = \bar{X}_1 - \bar{X}_2 \pm t_{1-\frac{\alpha}{2};n_1+n_2-2} \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

donde:

- \bar{X}_1 y \bar{X}_2 son los promedios de las muestras de las poblaciones 1 y 2.
 - n_1 y n_2 son los tamaños de las respectivas muestras.
 - σ_1^2 y σ_2^2 son las varianzas poblacionales.
 - S_1^2 y S_2^2 son las varianzas muestrales.
 - $\alpha = 1 - \text{NC}$. Es decir, si nos piden 95% de confianza, $\alpha = 1 - 0,95 = 0,05$.
 - $Z_{1-\frac{\alpha}{2}}$ es un fractil de la distribución normal estándar tal como lo estudiamos en el capítulo 6.
 - $t_{1-\frac{\alpha}{2};v}$ es un fractil de la distribución t-Student con v grados de libertad, tal como lo estudiamos en el capítulo 7.
- Los fractiles z y t se obtienen de las correspondientes tablas.

Problemas típicos

1) Se desea comparar el servicio de dos proveedores de internet. A tal efecto, se miden en varias ocasiones los tiempos que tardan en transmitir un paquete de 1 Mb. Las mediciones están expresadas en segundos.

• **Proveedor 1:** Se hicieron 9 mediciones. Se obtuvieron los valores 30, 33, 32, 34, 30, 23, 33, 31, 33.

• **Proveedor 2:** Se hicieron 11 mediciones. Se obtuvieron, en kb/s, los valores 26, 17, 22, 18, 21, 17, 23, 13, 24, 20, 19.

Estime mediante un intervalo del 90% confianza la diferencia entre los tiempos medios que tardan los dos proveedores en transmitir 1 Mb.

Resolución

El tiempo que tarda el proveedor 1 en transmitir 1 Mb es una variable aleatoria, con media μ_1 y varianza σ_1^2 . Análogamente, el tiempo que tarda el proveedor 2 en transmitir 1 Mb es una variable aleatoria, con media μ_2 y varianza σ_2^2 . Nos están pidiendo estimar la diferencia entre μ_1 y μ_2 , es decir, la diferencia entre las medias de los tiempos que tardan los proveedores en transmitir 1 Mb. Se pide estimar esa diferencia mediante un intervalo de confianza. Esto servirá para sacar una conclusión del estilo "el proveedor 2 tarda entre 2.5 y 4.1 segundos menos que el proveedor 1 en transmitir 1 Mb".

No conocemos las varianzas poblacionales σ_1^2 y σ_2^2 , por lo cual deberemos estimarlas con las varianzas muestrales S_1^2 y S_2^2 . Estamos en lo que hemos denominado **caso 2**.

Calculamos las medias muestrales:

- $\bar{X}_1 = 31$
- $\bar{X}_2 = 20$

Calculamos las varianzas muestrales:

- $S_1^2 = 11$
- $S_2^2 = 13.8$

Usamos las fórmulas. Los límites del intervalo vienen dados por:

$$L_{1;2} = X_1 - X_2 \pm t_{1-\frac{\alpha}{2};v} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 31 - 20 \pm t_{0.975;v} \sqrt{\frac{11}{9} + \frac{13.8}{11}} = 11 \pm 1.574 t_{0.95;v}$$

Los grados de libertad vienen dados por:

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}} = \frac{\left(\frac{11}{9} + \frac{13.8}{11}\right)^2}{\frac{\left(\frac{11}{9}\right)^2}{9-1} + \frac{\left(\frac{13.8}{11}\right)^2}{11-1}} = 17.83 \approx 18$$

Luego:

$$L_{1;2} = 11 \pm 1.574 t_{0.975;18} = 11 \pm 1.574 \cdot 1.7341$$

Es decir, obtenemos que $L_1 = 8.3$ y $L_2 = 13.7$.

La conclusión es: "El proveedor 2 es más rápido que el 1, y se puede afirmar con un 90% de confianza que el tiempo medio que habitualmente le toma al proveedor 1 transmitir 1 Mb es entre 8.3 y 13.7 segundos mayor que el tiempo que habitualmente le toma al proveedor 2 transmitir 1 Mb".

2) Se acaba de recibir la información de que el cableado y los routers utilizados por los proveedores estudiados son de calidad equivalente, y eso permite suponer que las varianzas de sus tasas de transferencia son iguales. Repita el análisis anterior, incorporando dicho conocimiento.

Resolución

Seguimos sin conocer las verdaderas varianzas, pero ahora nos informan que son iguales entre sí. Es decir, ahora estamos en el **caso 3**.

Usamos la fórmula:

$$L_{1;2} = 31 - 20 \pm t_{0.95;9+11-2} \sqrt{\frac{(9-1)11 + (11-1)13.8}{9+11-2} \left(\frac{1}{9} + \frac{1}{11}\right)} = 11 \pm t_{0.95;18} 1.28$$

Es decir, obtenemos que $L_1 = 8.8$ y $L_2 = 13.2$.

Vemos que el intervalo de confianza nos quedó más pequeño. Esto se debe a que en el segundo problema se tiene más información que en el primero, porque se **sabe** que las varianzas son iguales. Cuanta más información se tiene, más confianza se tiene, por eso para un mismo nivel de confianza del 90%, el intervalo construido con más información es más pequeño que el intervalo construido con menos

información.

3) Los dos proveedores que estábamos estudiando han informado las varianzas de sus respectivos tiempos de transferencia habituales de 1 Mb: el proveedor 1 ha informado que su varianza es 10, y el proveedor 2 ha informado que su varianza es 15. Repita el análisis.

Resolución

Ahora las varianzas poblacionales son conocidas, por lo cual ya no hay necesidad de aproximarlas mediante las muestrales. Estamos en lo que denominamos **caso 1**.

Se sabe que:

- $\sigma_1^2 = 10$

- $\sigma_2^2 = 15$

Usamos la fórmula:

$$L_{1;2} = \bar{X}_1 - \bar{X}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 31 - 20 \pm z_{0.95} \sqrt{\frac{10}{9} + \frac{15}{11}} = 11 \pm 1.645 \cdot 1.573$$

Es decir, obtenemos que $L_1 = 8.4$ y $L_2 = 13.6$.

Intervalo de confianza para la diferencia de dos proporciones

Sean las proporciones p_1 y p_2 . Se desea construir el intervalo de confianza de la diferencia $p_1 - p_2$, para lo cual se toman de cada muestras de tamaños n_1 y n_2 respectivamente, contabilizándose X_1 y X_2 resultados favorables respectivamente.

Se estimarán p_1 y p_2 mediante $\hat{p}_1 = X_1 / n_1$ y $\hat{p}_2 = X_2 / n_2$.

Se requiere que n_1 y n_2 sean lo suficientemente grandes para que se cumplan:

$$n_1 \hat{p}_1 \geq 5 ; n_1 (1 - \hat{p}_1) \geq 5 ; n_2 \hat{p}_2 \geq 5 ; n_2 (1 - \hat{p}_2) \geq 5$$

Los límites L_1 y L_2 del intervalo de confianza son:

$$L_{1,2} = \hat{p}_1 - \hat{p}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

donde:

$$\hat{p}_1 = \frac{X_1}{n_1} \quad ; \quad \hat{p}_2 = \frac{X_2}{n_2}$$

- n_1 y n_2 son los tamaños de las muestras.
- X_1 y X_2 las cantidades de éxitos en esas muestras.
- $\alpha = 1 - \text{NC}$. Es decir, si nos piden 95% de confianza, $\alpha = 1 - 0,95 = 0,05$.

• $z_{1-\frac{\alpha}{2}}$ es un fractil de la normal estándar tal como lo estudiamos en el capítulo 6.

Problemas típicos

1) Se cree que en la provincia de Santa Fe hay más gente con ojos claros que en la provincia de Entre Ríos. Para eso se toma una muestra de 200 personas de Sante Fe y 500 de Entre Ríos, obteniéndose 70 y 120 personas con ojos claros respectivamente.

Indique un intervalo de confianza del 95% para la diferencia entre las proporciones de personas con ojos claros en ambas provincias.

Resolución

Las proporciones muestrales son:

$$\hat{p}_S = \frac{X_S}{n_S} = \frac{70}{200} = 0.35 \quad ; \quad \hat{p}_E = \frac{X_E}{n_E} = \frac{120}{500} = 0.24$$

Usamos la fórmula:

$$L_{1;2} = 0.35 - 0.24 \pm z_{0.975} \sqrt{\frac{0.35(1-0.35)}{200} + \frac{0.24(1-0.24)}{500}} = 0.11 \pm 1.96 \cdot 0.03876$$

Obtuvimos que: $L_1 = 0.034$ y $L_2 = 0.186$

Esto quiere decir que se puede afirmar con un 95% de confianza que en la provincia de Santa Fe hay entre 3.4 y 18.6 más puntos porcentuales de personas con ojos claros que en la provincia de Entre Ríos.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 22 de julio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 12 de julio de 2004

CAPÍTULO X

Pruebas de hipótesis

(o "ensayos de hipótesis" o "tests de hipótesis")

En los capítulos anteriores estudiamos cómo estimar parámetros y cómo calcular la confianza que se le puede tener a los estimadores empleados para tal fin. Ahora estudiaremos cómo, en base a datos estadísticos, decidir si se deben aceptar o rechazar determinadas hipótesis acerca de los parámetros estimados.

Más precisamente, tomaremos dos hipótesis mutuamente excluyentes H_0 y H_A ("hipótesis nula" e "hipótesis alternativa") y analizaremos si en base a lo que observamos en la muestra, se puede o no rechazar la hipótesis nula H_0 frente a la hipótesis alternativa H_A .

Por ejemplo, si la hipótesis H_0 dice "la media de la población es 10" y la hipótesis alternativa dice "la media de la población es mayor que 10", analizaremos si debemos rechazar H_0 frente a la hipótesis alternativa. Es decir, si en base a los datos que obtenemos de la muestra (y a los riesgos que estamos dispuestos a correr) es más **razonable** H_A que H_0 . Por otra parte, el hecho de no rechazar H_0 no implicará necesariamente aceptarla, porque el hecho de aceptarla también conlleva un cierto nivel de riesgo.

Justamente, al igual que para definir un intervalo de confianza teníamos que adoptar un nivel de confianza, para efectuar un ensayo de hipótesis debemos adoptar un determinado nivel de riesgo. Antes de definir los riesgos, veamos cuáles son los dos errores posibles que podríamos cometer al tomar la decisión de rechazar o no rechazar H_0 :

- Rechazar H_0 cuando en realidad era verdadera. (Error tipo I).
- No rechazar H_0 cuando en realidad era falsa. (Error tipo II).

Así como el nivel de confianza α que tomamos para un intervalo de confianza es la probabilidad de que el intervalo contenga al verdadero valor, los riesgos que corremos en el ensayo de hipótesis también son probabilidades:

- Llamamos α a la probabilidad de cometer el error tipo I.
- Llamamos β a la probabilidad de cometer el error tipo II.

Nos referiremos a α como "nivel de significación" y a $1-\beta$ como "potencia de la prueba".

La forma de decidir es:

- 1) Se elige un estimador para el parámetro desconocido (por ejemplo, para la media de una población se usa \bar{X} , para una proporción se usa \hat{p} , para la varianza se usa S^2 , etc.).
- 2) Teniendo en cuenta las características del estimador empleado y los errores que estamos dispuestos a cometer, se determina, entre todos los valores posibles que podrá arrojar el estimador al hacer el experimento, el conjunto de valores del estimador para los cuales rechazaríamos H_0 . Dicho conjunto de valores se denomina región crítica (RC).
- 3) Se toma la muestra y se calcula el valor del estimador.
- 4) Si el valor que arrojó el estimador pertenece a la región crítica, se rechaza H_0 .

Hasta aquí hemos dado las características comunes a todos los ensayos de hipótesis. En adelante haremos un estudio sobre los ensayos que involucran a las hipótesis más habituales. Sin embargo, el problema típico número 4 de esta sección constituye un ejemplo de ensayo con hipótesis que no son las comunes que estudiaremos a continuación.

Tipos de hipótesis más comunes

Los 4 tipos de hipótesis más comunes son:

- por igual (Ej. $\mu = 40$)
- por mayor (Ej. $\mu > 40$ ó $\mu \geq 40$)
- por menor (Ej. $\mu < 40$ ó $\mu \leq 40$)
- por distinto (Ej. $\mu \neq 40$)

En adelante nuestro estudio será sobre estos tipos de hipótesis.

Elección de H_0 (hipótesis nula)

La decisión de cuál de las dos hipótesis será elegida como H_0 depende de cuáles sean los tipos de hipótesis involucradas. Dadas dos hipótesis entre las cuatro más comunes, la forma de elegir cuál de las dos hipótesis se elige como H_0 es la siguiente:

- Cuando una de las hipótesis es por igual, entonces esa se elige como H_0 .
- Si las dos son por igual, se elige como H_0 la que más "lejos" tenga al estimador.

(Ej.: si las hipótesis son " $p = 30$ " y " $p = 50$ ", y tenemos que $\hat{p} = 45$, elegiremos como H_0 a la hipótesis " $p = 30$ ".

- Si una es por menor y la otra por mayor, se elige H_0 con el mismo criterio que en el punto anterior: se elige como H_0 la que más "lejos" tenga al estimador. Luego, la que sea elegida como H_0 será tratada a los fines prácticos como por igual en vez de por mayor o por menor.

- Cualquier otra combinación de los 4 tipos de hipótesis comunes dados no es posible, debido a que las dos hipótesis no resultarían mutuamente excluyentes como se requiere.

Tipos de pruebas comunes

Por lo tanto, quedan determinados 4 tipos de pruebas:

- Igual contra mayor
 - $H_0: \theta = \theta_0$
 - $H_A: \theta > \theta_0$
- Igual contra menor
 - $H_0: \theta = \theta_0$
 - $H_A: \theta < \theta_0$
- Igual contra distinto
 - $H_0: \theta = \theta_0$
 - $H_A: \theta \neq \theta_0$
- Igual contra igual
 - $H_0: \theta = \theta_0$
 - $H_A: \theta = \theta_1$

Rechazo o no rechazo de H_0

Las decisiones en sí se toman de la siguiente manera:

Primero se elige un estimador $\hat{\theta}$ para el parámetro en cuestión. Por ejemplo, para la media de una población normal se usa \bar{X} , para una proporción se usa \hat{p} , para la varianza se usa S^2 , etc.

Luego se determina cuáles son los valores del estimador para los cuales rechazaremos H_0 . Esto depende del tipo de prueba:

- igual contra mayor: se encontrará un valor θ_c (crítico) tal que H_0 se rechaza si el estimador resulta mayor que θ_c . Esto es lógico porque lo que estamos analizando es hasta qué valor del estimador seguiremos pensando que es más razonable decir que el parámetro es igual a un valor contra la alternativa de que en realidad sea mayor que dicho valor. Como ejemplo podemos pensar:

$$H_0: \mu = 5$$

$$H_A: \mu > 5$$

donde μ es la media de una población, con lo cual el estimador utilizado será \bar{X} . Si el valor crítico que determinamos fuera por ejemplo $X_c = 7$, entonces tomamos el promedio de la muestra, y si es mayor que 7 rechazamos H_0 , y si es menor que 7 no la rechazamos.

El conjunto de los valores posibles del estimador (en este caso el promedio) para

los cuales rechazamos H_0 se denomina "región crítica" o "región de rechazo". Su complemento se denomina "región de aceptación". Se sigue que $RC \cup RA$ es el conjunto de todos los valores posibles del estimador.

- igual contra menor: análogamente, encontraremos un θ_c tal que H_0 se rechaza si el estimador resulta menor que dicho valor.
- igual contra distinto: en este caso hay que determinar dos valores alrededor del valor propuesto por H_0 , tales que si el estimador no cae entre ellos, rechazaremos H_0 porque el estimador cae demasiado lejos del valor que propone H_0 como para considerar que la igualdad es más razonable que la desigualdad. Recomendamos dedicar unos momentos a comprender esto. Luego se verá la interpretación gráfica.
- igual contra igual: se determinará un θ_c que se encuentre en algún lugar en el medio entre los valores propuestos por H_0 y H_A , y se rechazará H_0 si el estimador cae del lado del valor propuesto por H_A .

Determinación de los θ_c y la región crítica

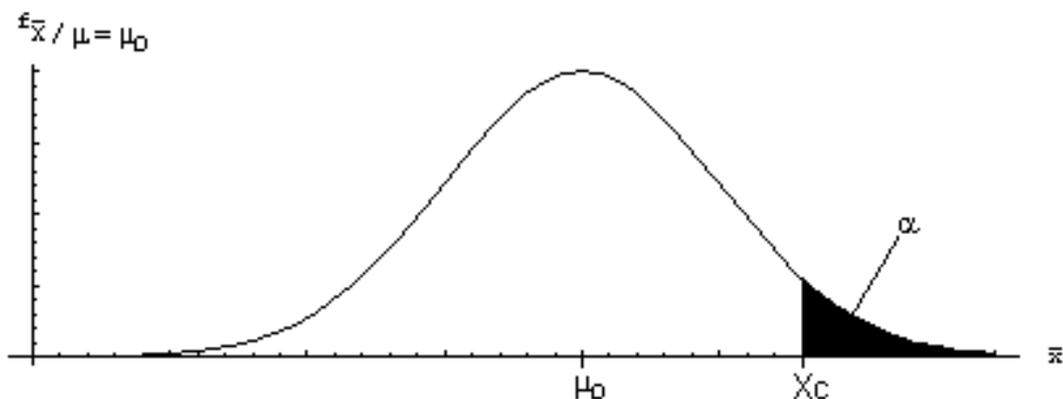
¿Cómo se determina el valor de θ_c que debe usarse?

Si H_0 es verdadera, la probabilidad de que la rechazemos es α (la probabilidad de cometer el error tipo I). Y rechazamos H_0 cuando el estimador cae en la región crítica. Entonces la probabilidad de que el estimador caiga en la región crítica, siendo H_0 verdadera, debe ser α . Por lo tanto, conociendo la distribución de H_0 , podemos determinar el θ_c tal que el área encerrada sobre la región crítica sea α . Más adelante veremos la forma metódica de hacerlo.

Interpretación gráfica

Por ejemplo, en el caso de la estimación de la media μ mediante el promedio muestral \bar{X} , los X_c se eligen de la siguiente manera:

- Para el caso de que la prueba sea por igual contra mayor:



Se rechaza $H_0: \mu = \mu_0$ si \bar{X} resulta mayor que X_c , donde X_c es tal que el área de campana centrada en μ_0 encerrada a la derecha de X_c es α . ¿Qué significa esto? Dijimos que α es la probabilidad de cometer el error tipo I (rechazar H_0 cuando en realidad H_0 es verdadera). Luego:

$$P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = \alpha$$

Como la distribución de \bar{X} es $N\left(\mu; \sigma/\sqrt{n}\right)$, entonces decir que H_0 es verdadera es decir que $\mu = \mu_0$, lo cual es lo mismo que decir que la distribución de \bar{X} , dado que

H_0 es verdadera, es $N\left(\mu_0; \sigma/\sqrt{n}\right)$. Luego decir que H_0 es verdadera es lo mismo que decir que la distribución de \bar{X} es la campana centrada en μ_0 que vemos en la ilustración.

Entonces asumiendo que H_0 es verdadera (o sea que la distribución de \bar{X} es la campana que dibujamos), la probabilidad condicional que teníamos antes se reduce a:

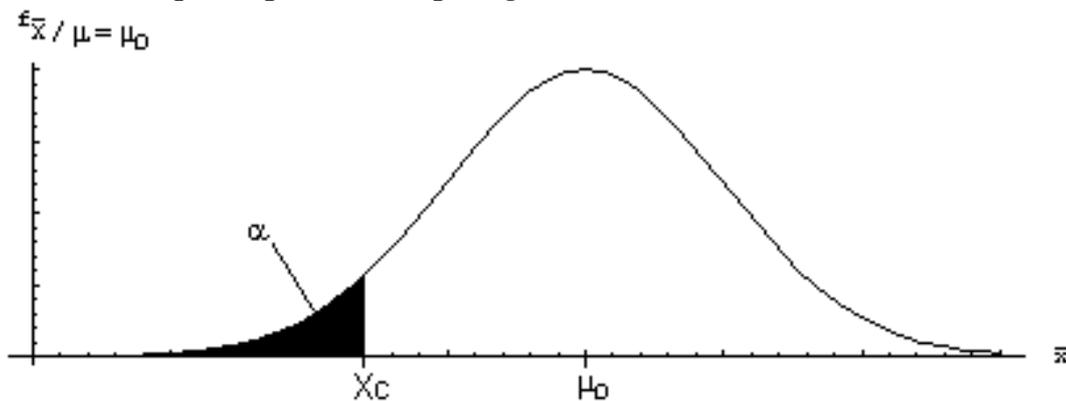
$$P(\text{rechazar } H_0) = \alpha$$

Y habíamos dicho que rechazamos H_0 si $\bar{X} > X_c$. Luego, la probabilidad queda:

$$P(\bar{X} > X_c) = \alpha$$

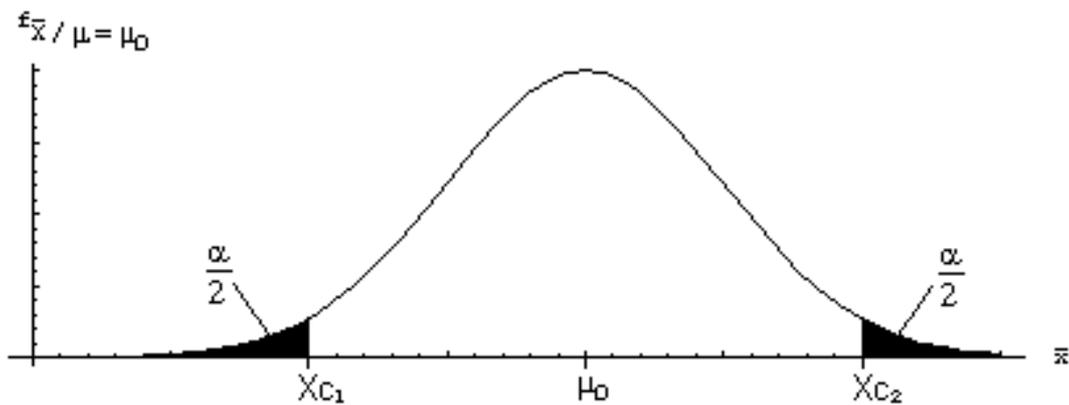
Con lo cual finalmente entendemos por qué X_c es tal que el área bajo la campana centrada en μ_0 encerrada a la derecha de X_c es α .

- Para el caso de que la prueba sea por igual contra menor:



Este caso es análogo al anterior. Rechazamos H_0 si \bar{X} resulta menor que X_c , donde X_c es tal que el área de campana centrada en μ_0 encerrada a la izquierda de X_c es α .

- Para el caso de que la prueba sea por igual contra distinto:



En el caso de igual contra distinto, se rechaza H_0 si \bar{X} resulta menor que X_{C1} o mayor que X_{C2} . Vemos que la probabilidad α se reparte entre las dos colas.

- Para el caso de que la prueba sea por igual contra igual, el gráfico queda similar al presentado para la prueba por igual contra mayor (si el valor propuesto por H_0 es menor al propuesto por H_A), y similar al presentado para la prueba por igual contra menor (si el valor propuesto por H_0 es mayor al propuesto por H_A).

Pruebas a una cola y a dos colas

Vemos que en las pruebas contra mayor y contra menor, estamos tomando la probabilidad de una sola cola de la normal, y que en la prueba contra distinto estamos tomando la probabilidad de dos colas de la normal. Es por eso que a veces se habla de pruebas "a una cola" y "a dos colas". La prueba "a dos colas" es la contra distinto. Las demás son "a una cola".

Potencia del test

Como se dijo antes, la potencia un test es $1 - \beta$. Es decir, 1 - la probabilidad de no rechazar H_0 cuando en realidad es falsa. Esto es, la probabilidad de rechazar H_0 cuando es falsa. Es por eso que se la llama potencia.

En las subsecciones donde se justifican las fórmulas que usaremos en cada tipo de prueba para cada parámetro, veremos cómo se calcula la potencia de un determinado test.

Normalmente deseamos que la potencia, es decir, la probabilidad de rechazar H_0 cuando es falsa, sea alta. Por eso puede suceder que para una determinada prueba, no estemos conformes con el valor que alcanza la potencia. En ese caso, hay dos formas de mejorarla:

- Aumentando el tamaño de la muestra.
- Aumentando α (es decir, sacrificando un poco el nivel de significación).

Para una interpretación gráfica, se puede ver el gráfico proporcionado para α y β en la deducción de las fórmulas para las pruebas sobre medias, de la sección 10.2 (pruebas sobre medias).

Método para diseñar ensayos de hipótesis

Para comprender este método se recomienda leer detenidamente todo lo dicho hasta ahora en este capítulo, ya que en cada una de las subsecciones se indica cómo realizar lo que pide cada paso del método.

Los pasos son los siguientes:

1) Determinar α , H_0 , H_A , y el estimador que se usará para la prueba. Puede ser que además conozcamos β para algún valor en particular.

2) En función de H_0 y H_A , determinar la forma que tendrá la región crítica (sin calcular los θ_c)

3) Encontrar la distribución del estimador, en función del parámetro desconocido.

4) Plantear:

$$P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = \alpha$$

5) Reescribir el suceso "se rechaza H_0 " dejándolo en función del estimador y el valor crítico. (Por ejemplo, $\bar{X} > X_c$, $\hat{p} < p_c$, $\hat{\theta} < \theta_{c1} \wedge \hat{\theta} > \theta_{c2}$, etc.).

6) Especializar la distribución del estimador obtenida en 3) en el valor del parámetro propuesto por H_0 . Asumiendo esto, la probabilidad condicional que tenemos igualada a α deja de ser condicional.

7) De la ecuación resultante, despejar el (o los) valores críticos, con lo cual ya tenemos la regla de decisión.

8) Si la expresión hallada en 7 queda en función de n o algún otro elemento desconocido, necesitamos conocer β para algún valor, y despejaremos el valor del elemento desconocido igualando la potencia a $1-\beta$.

Si se desea o necesita estudiar la potencia:

9) La potencia es $1-\beta = 1 - P\left(\text{no rechazar } H_0 / H_0 \text{ falsa}\right)$

10) Reescribir el suceso "no se rechaza H_0 " dejándolo en función del estimador y el valor crítico (debería quedar el complemento de lo obtenido en 5).

11) Especializar la distribución del estimador obtenida en 3) en un valor o expresión que haga falsa a H_0 . Dicho valor puede ser dato o podemos tener que proponerlo. Asumiendo esto, la probabilidad condicional deja de serlo.

12) Despejar y/o calcular lo que sea necesario.

Nota: este método es bastante general. En las siguientes secciones de este capítulo se estudiará su particularización a los siguientes casos:

- pruebas sobre la media de una población normal y/o con muestra grande.
- pruebas sobre la varianza de una población normal y/o con muestra grande.

- pruebas sobre proporciones con muestras grandes.
- pruebas sobre la diferencia de dos medias.
- pruebas sobre la diferencia de dos proporciones.
- pruebas sobre la comparación de dos varianzas desconocidas.

y se enunciarán las fórmulas y reglas de decisión para dichos casos. Para otros casos, se deberá usar el método más general que acabamos de enunciar.

Resumen del método para las hipótesis más comunes

Tipo de prueba	Hipótesis	Se rechaza H_0 si:	Los valores críticos se despejan de:	Dado θ_v , la potencia ($1 - \beta$) es:
igual contra mayor	$H_0: \theta = \theta_0$ $H_A: \theta > \theta_0$	$\hat{\theta} > \theta_c$	$P\left(\hat{\theta} > \theta_c / \theta = \theta_0\right) = \alpha$	$1 - P\left(\hat{\theta} \leq \theta_c / \theta = \theta_v\right)$
igual contra menor	$H_0: \theta = \theta_0$ $H_A: \theta < \theta_0$	$\hat{\theta} < \theta_c$	$P\left(\hat{\theta} < \theta_c / \theta = \theta_0\right) = \alpha$	$1 - P\left(\hat{\theta} \geq \theta_c / \theta = \theta_v\right)$
igual contra distinto	$H_0: \theta = \theta_0$ $H_A: \theta \neq \theta_0$	$\hat{\theta} < \theta_{c1}$ o bien $\hat{\theta} > \theta_{c2}$	$P\left(\theta_{c1} < \hat{\theta} < \theta_{c2} / \theta = \theta_0\right) = 1 - \alpha$	$1 - P\left(\theta_{c1} \leq \hat{\theta} \leq \theta_{c2} / \theta = \theta_v\right)$
igual contra igual	$H_0: \theta = \theta_0$ $H_A: \theta = \theta_1$	Según corresponda: • si $\theta_0 < \theta_1$: como si fuera igual contra mayor • si $\theta_0 > \theta_1$: como si fuera igual contra menor		

Problemas típicos

1) Mientras esperan la luz verde del semáforo para poder cruzar la calle, dos estudiantes de ingeniería discuten acerca de la distribución de la cantidad de autos que circulan por la mencionada calle. Ambos están de acuerdo en que los autos pasan según un proceso Poisson, pero el estudiante A opina que la media del proceso es 15 autos por minuto y el estudiante B sostiene que dicha media es mayor que 15 autos por minuto. Abandonan sus planes de cruzar la calle, y durante 1 minuto cuentan la cantidad de autos que pasan, obteniendo que en ese minuto pasaron 17 autos. A un nivel de significación del 5% decida cuál de los dos estudiantes tiene razón. ¿Cuál sería la potencia del test si la verdadera media fuera 25?

Resolución

El nivel de significación es $\alpha = 0,05$.

Es una prueba de igual contra mayor, cuyas hipótesis son:

$$H_0: \mu = 15$$

$$H_A: \mu > 15$$

El estimador que usaremos será la cantidad de autos que pasaron en el minuto observado, y lo llamaremos X .

Por ser una prueba de igual contra mayor, rechazaremos H_0 si X es mayor a un cierto X_c . Es decir, la región crítica o de rechazo estará constituida por todos los valores de X mayores a X_c .

La distribución de X es una Poisson con media μ (μ es el parámetro desconocido que estamos estimando). $X: \text{Pois}(\mu)$.

La probabilidad de cometer el error tipo I es $\alpha = 0.05$, con lo cual:

$$P\left(\frac{\text{rechazar } H_0}{H_0 \text{ verdadera}}\right) = \alpha$$

Rechazamos H_0 cuando $X > X_c$. Por lo tanto podemos reescribir:

$$P\left(\frac{X > X_c}{H_0 \text{ verdadera}}\right) = \alpha$$

Que H_0 sea verdadera significa que $\mu = 15$, con lo cual queda:

$$P\left(\frac{X > X_c}{\mu = 15}\right) = \alpha$$

Si condicionamos la distribución de X a que μ valga 15, queda $X: \text{Pois}(15)$ y podemos escribir:

$$P(X > X_c) = \alpha$$

Luego como X tiene una distribución de Poisson, sabemos que sus probabilidades se distribuyen según:

$$P(X=x) = \frac{e^{-\mu} \mu^x}{x!}$$

Con lo cual estamos en condiciones de averiguar cuánto tiene que valer X_c para que se cumpla que

$$P(X > X_c) = \alpha = 0,05$$

Invertiremos la inecuación para poder calcularla:

$$P(X \leq X_c) = 1 - \alpha$$

Buscamos X_c tal que:

$$P(X \leq X_c) = 0,95$$

$$\text{Si } X_c \text{ valiera } 0, \quad P(X \leq X_c) = P(X \leq 0) = P(X = 0) = \frac{e^{-15} 15^0}{0!} = 0,0000003$$

$$\text{Si } X_c \text{ valiera } 1, \quad P(X \leq X_c) = P(X \leq 1) = P(X = 0) + P(X = 1) = 0,0000049$$

$$\text{Si } X_c \text{ valiera } 2, \quad P(X \leq X_c) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0,0000393$$

$$P(X \leq X_c) \approx 0,95$$

Continuamos, hasta que para $X_c = 21$ obtenemos

Con lo cual como $X < X_c$, no rechazamos H_0 . Por lo tanto, en principio le damos la razón al estudiante A, porque a la luz de lo observado, resulta más razonable pensar que $\mu = 15$ y no que $\mu > 15$. Es decir, no hay suficiente evidencia estadística como

para rechazar que μ vale 15.

La potencia es $1 - \beta = 1 - P\left(\text{no rechazar } H_0 / H_0 \text{ falsa}\right)$

No rechazamos H_0 cuando $X \leq X_c$. Reescribimos:

$$\text{Potencia} = 1 - P\left(X \leq X_c / H_0 \text{ falsa}\right)$$

Estamos suponiendo el caso hipotético de que el verdadero valor de μ fuera 25, con lo cual:

$$\text{Potencia} = 1 - P\left(X \leq X_c / \mu = 25\right)$$

Si condicionamos la distribución de X a que μ valga 25, queda $X:\text{Pois}(25)$ y podemos escribir:

$$\text{Potencia} = 1 - P(X \leq X_c)$$

Y ahora calculamos la potencia:

$$1 - P(X \leq X_c) = 1 - P(X \leq 21) = 1 - (P(X=0) + P(X=1) + \dots + P(X=21))$$

Y como $X:\text{Pois}(25)$, haciendo la cuenta resulta:

Potencia = 0.75. Como la potencia para un valor razonablemente mayor que 15 resultó razonablemente cercana a 1, el test es razonablemente bueno, con lo cual confirmamos la decisión de darle la razón al estudiante A.

2) Para una distribución Poisson, se desea ensayar la hipótesis nula $\mu = 6$ contra la hipótesis alternativa $\mu < 6$. Para eso se tomó una muestra de tamaño 5, que arrojó los siguientes valores: 3, 7, 1, 6, 3. Tome una decisión, a un nivel de significación del 10%.

Resolución

El nivel de significación es $\alpha = 0,1$.

Es una prueba de igual contra menor, cuyas hipótesis son:

$$H_0: \mu = 6$$

$$H_A: \mu < 6$$

El estimador que usaremos es \bar{X} , el promedio de la muestra. Si la variable aleatoria

$$\bar{X} = \frac{Y}{5}$$

Y es la suma de los elementos de la muestra, entonces

Por ser una prueba de igual contra menor, rechazaremos H_0 si \bar{X} es menor a un cierto X_c . Es decir, la región crítica o de rechazo estará constituida por todos los valores de \bar{X} menores a X_c .

La variable aleatoria Y es la suma de 5 variables Poisson independientes con media μ , con lo cual $Y:\text{Pois}(5\mu)$.

$$P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = \alpha$$

Rechazamos H_0 cuando $\bar{X} < X_c$. Por lo tanto podemos reescribir:

$$P\left(\bar{X} < X_c \middle/ H_0 \text{ verdadera}\right) = \alpha$$

Que H_0 sea verdadera significa que $\mu = 6$, con lo cual queda:

$$P\left(\bar{X} < X_c \middle/ \mu = 6\right) = \alpha$$

Reemplazamos por la definición de \bar{X} en términos de Y y n .

$$P\left(\frac{Y}{5} < X_c \middle/ \mu = 6\right) = P\left(Y < 5X_c \middle/ \mu = 6\right) = \alpha$$

Si condicionamos la distribución de Y a que μ valga 6, queda $Y: \text{Pois}(30)$ y podemos escribir:

$$P(Y < 5X_c) = \alpha$$

Luego como Y tiene una distribución de Poisson, sabemos que sus probabilidades se distribuyen según:

$$P(Y=y) = \frac{e^{-\mu} \mu^y}{y!}$$

Con lo cual estamos en condiciones de averiguar cuánto tiene que valer nX_c para

que se cumpla que $P(Y < 5X_c) = \alpha = 0,1$

Sumando las $P(Y = y)$ encontramos que:

$$P(Y < 23) = 0.08057$$

$$P(Y < 24) = 0.11465$$

Interpolando, podríamos decir que si $y = 23.5$, entonces $P(Y < y) \cong 0.1$.

(En realidad el valor 23.5 es imposible porque la distribución Poisson tiene probabilidad no nula solamente para el número 0 y los números naturales).

Entonces llegamos a que $5X_c \cong 23.5$, con lo cual $X_c \cong 4.7$.

Hemos encontrado X_c . Enunciamos la regla de decisión: rechazaremos H_0 si $\bar{X} < 4.7$.

Si calculamos \bar{X} para la muestra tomada obtenemos $\bar{X} = 4.2$, con lo cual rechazamos " $\mu = 6$ ", a un nivel de significación del 10%.

3) La duración de un determinado componente tiene una distribución exponencial negativa, pero no se sabe si la media de dicha distribución es 2 ó 5. Para decidirlo, se toma un componente y se mide su duración. Si resulta mayor que 3, se decide que la media vale 5, y si resulta menor que 3, se decide que la media vale 2. ¿Cuáles son las probabilidades de cometer los errores tipo I y tipo II utilizando dicha regla de decisión?

Resolución

Estamos ante una prueba de igual contra igual, en la cual:

$$H_0: \mu = 2$$

$$H_A: \mu = 5$$

Nos dicen que se toma como estimador de la media la duración del componente observado. Denominaremos a dicho estimador X , y su distribución es $X: \text{ExpNeg}(\lambda)$ donde por propiedades de la distribución exponencial negativa, $\lambda = 1/\mu$.

Normalmente, usaríamos α para buscar un valor crítico tal que si el estimador es mayor que dicho valor crítico, rechazamos H_0 y si es menor, no rechazamos H_0 .

Este problema es al revés: nos dicen el valor crítico y nos piden el α .

$$\alpha = P\left(\begin{array}{c} \text{rechazar } H_0 \\ H_0 \text{ verdadera} \end{array}\right) = P\left(\begin{array}{c} X > X_c \\ H_0 \text{ verdadera} \end{array}\right) = P\left(\begin{array}{c} X > X_c \\ \mu = 2 \end{array}\right)$$

Si condicionamos la distribución de X a que $\mu = 2$, queda: $X: \text{ExpNeg}(1/2)$ y luego:

$$P\left(\begin{array}{c} X > X_c \\ \mu = 2 \end{array}\right) = P\left(\begin{array}{c} X > X_c \\ \lambda = 1/2 \end{array}\right) = P\left(\begin{array}{c} X > 2 \\ \lambda = 1/2 \end{array}\right) = \int_2^{\infty} \frac{1}{2} e^{-\frac{1}{2}x} dx = 0,37$$

Luego $\alpha = 0,37$.

$$\beta = P\left(\begin{array}{c} \text{no rechazar } H_0 \\ H_0 \text{ falsa} \end{array}\right)$$

No rechazamos H_0 cuando $X < X_c$. Reescribimos:

$$\beta = P\left(\begin{array}{c} X < X_c \\ H_0 \text{ falsa} \end{array}\right)$$

Si H_0 es falsa, entonces $\mu \neq 2$. La única otra posibilidad que las hipótesis permiten es que $\mu = 5$. Reescribimos:

$$\beta = P\left(\begin{array}{c} X < X_c \\ \mu = 5 \end{array}\right)$$

Si condicionamos la distribución de X a que $\mu = 5$, queda: $X: \text{ExpNeg}(1/5)$ y luego:

$$P\left(\begin{array}{c} X < X_c \\ \mu = 5 \end{array}\right) = P\left(\begin{array}{c} X < X_c \\ \lambda = 1/5 \end{array}\right) = P\left(\begin{array}{c} X < 2 \\ \lambda = 1/5 \end{array}\right) = \int_0^2 \frac{1}{5} e^{-\frac{1}{5}x} dx = 0,33$$

Luego $\beta = 0,33$.

4) Las normas exigen que las máquinas que producen un determinado tipo de pieza estén calibradas para que la media del peso esté entre 10 y 12 gramos. Se desea verificar si una de las máquinas cumple con los requerimientos. Se sabe que el peso de las piezas producidas por esa máquina es una variable normal con $\sigma = 6$. Se toma una muestra de 16 piezas producidas por esa máquina, obteniéndose $\bar{X} = 9$. Decida a un nivel de significación del 5%.

Resolución

Este ejemplo está dado con la intención de mostrar un test cuyas hipótesis no son las habituales. Las hipótesis que tendremos son:

$$H_0: \mu \in [10;12]$$

$$H_A: \mu \notin [10;12]$$

El procedimiento consistirá en tomar los límites L_1 y L_2 , tales que si \bar{X} cae dentro de esos límites, consideraremos más razonable H_0 , y en caso contrario consideraremos más razonable H_A . ¿Por qué no tomamos como límites directamente el 10 y el 12? Porque queremos fijar el riesgo en un 5%, y no sabemos cuál es el riesgo de tomar como límites el 10 y el 12.

El valor $\alpha = 0.05$ es la probabilidad de rechazar H_0 siendo esta verdadera. En este caso, 0.05 será la probabilidad de que \bar{X} caiga fuera de los límites L_1 y L_2 en caso de que la media efectivamente esté entre 10 y 12.

Para dividir equitativamente, tomaremos $\alpha/2$ como la probabilidad de que $\bar{X} < L_1$ aunque $\mu \in [10;12]$, y el otro $\alpha/2$ como la probabilidad de que $\bar{X} > L_2$ aunque $\mu \in [10;12]$. Es decir:

$$P\left(\bar{X} < L_1 / \mu \in [10;12]\right) = \frac{\alpha}{2}$$

$$P\left(\bar{X} > L_2 / \mu \in [10;12]\right) = \frac{\alpha}{2}$$

Si algo se cumple para el peor caso posible, debe también cumplirse para los casos menos extremos. En la primera probabilidad planteada, el peor caso posible es $\mu = 10$, porque cuanto más chico sea el verdadero valor de μ , mayor será la probabilidad de que $\bar{X} < L_1$. Análogamente, en la segunda probabilidad planteada, el peor caso posible es $\mu = 12$, porque cuanto más grande sea el valor de μ , mayor será la probabilidad de que $\bar{X} > L_2$.

Luego, escribimos:

$$P\left(\bar{X} < L_1 / \mu = 10\right) = \frac{\alpha}{2}$$

$$P\left(\bar{X} > L_2 / \mu = 12\right) = \frac{\alpha}{2}$$

Como se vio en capítulos anteriores, $\bar{X} \sim N\left(\mu ; \frac{\sigma}{\sqrt{n}}\right)$. En este caso, como $\sigma = 6$ y $n = 16$, queda $\bar{X} \sim N(\mu ; 1.5)$. Luego, podemos despejar L_1 y L_2 :

$$P\left(\bar{X} < L_1 / \mu = 10\right) = \Phi\left(\frac{L_1 - 10}{1.5}\right) = \frac{\alpha}{2} \Rightarrow \frac{L_1 - 10}{1.5} = z_{\frac{\alpha}{2}} \Rightarrow L_1 = 10 + 1.5 z_{\frac{\alpha}{2}}$$

$$P\left(\bar{X} > L_2 / \mu = 12\right) = 1 - \Phi\left(\frac{L_2 - 12}{1.5}\right) = \frac{\alpha}{2} \Rightarrow \frac{L_2 - 12}{1.5} = z_{1 - \frac{\alpha}{2}} \Rightarrow L_2 = 12 + 1.5 z_{1 - \frac{\alpha}{2}}$$

Obtenemos entonces:

$$L_1 = 7.06$$

$$L_2 = 14.94$$

Luego, como $\bar{X} = 9$ queda dentro de los límites encontrados, decidimos que es más razonable suponer que μ se encuentra entre 10 y 12 que suponer que no se encuentra entre 10 y 12.

Este tipo de test no es muy común debido a que, como habitualmente se conoce el valor de \bar{X} antes de diseñar el test, se hace directamente un test contra mayor o

contra menor según corresponda. Sin embargo, dicho proceder puede llegar a ser conceptualmente objetable, porque de esa forma los valores de la muestra estarían condicionando el tipo de test que se usará con ella.

Pruebas de hipótesis para la media de una población

La población debe ser normal y/o la muestra debe ser grande.

Se deben elegir H_0 y H_A como se indicó en la sección anterior, y luego:

Hipótesis	Desvío poblacional	Se rechaza H_0 si:	La potencia del test $(1-\beta)$ es:
$H_0: \mu = \mu_0$ $H_A: \mu > \mu_0$ "igual contra mayor"	conocido	$\bar{X} > \mu_0 + \frac{z_{1-\alpha} \sigma}{\sqrt{n}}$	$\Phi \left(\frac{\mu_v - X_c}{\sigma / \sqrt{n}} \right)$
	desconocido	$\bar{X} > \mu_0 + \frac{t_{1-\alpha; n-1} S}{\sqrt{n}}$	$\tau_{n-1} \left(\frac{\mu_v - X_c}{S / \sqrt{n}} \right)$
$H_0: \mu = \mu_0$ $H_A: \mu < \mu_0$ "igual contra menor"	conocido	$\bar{X} < \mu_0 - \frac{z_{1-\alpha} \sigma}{\sqrt{n}}$	$\Phi \left(\frac{X_c - \mu_v}{\sigma / \sqrt{n}} \right)$
	desconocido	$\bar{X} < \mu_0 - \frac{t_{1-\alpha; n-1} S}{\sqrt{n}}$	$\tau_{n-1} \left(\frac{X_c - \mu_v}{S / \sqrt{n}} \right)$
$H_0: \mu = \mu_0$ $H_A: \mu \neq \mu_0$ "igual contra distinto"	conocido	$\bar{X} > \mu_0 + \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$ o bien: $\bar{X} < \mu_0 - \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$	$1 - \Phi \left(\frac{X_{c2} - \mu_v}{\sigma / \sqrt{n}} \right) + \Phi \left(\frac{X_{c1} - \mu_v}{\sigma / \sqrt{n}} \right)$
	desconocido	$\bar{X} > \mu_0 + \frac{t_{1-\frac{\alpha}{2}; n-1} \sigma}{\sqrt{n}}$ o bien: $\bar{X} < \mu_0 - \frac{t_{1-\frac{\alpha}{2}; n-1} \sigma}{\sqrt{n}}$	$1 - \tau_{n-1} \left(\frac{X_{c2} - \mu_v}{S / \sqrt{n}} \right) + \tau_{n-1} \left(\frac{X_{c1} - \mu_v}{S / \sqrt{n}} \right)$
$H_0: \mu = \mu_0$ $H_A: \mu = \mu_1$ "igual contra igual"	si $\mu_0 < \mu_1$, los criterios coinciden con los usados para igual contra mayor		$\Phi \left(\frac{\mu_1 - X_c}{\sigma / \sqrt{n}} \right)$ o bien $\tau_{n-1} \left(\frac{\mu_1 - X_c}{S / \sqrt{n}} \right)$
	si $\mu_0 > \mu_1$, los criterios coinciden con los usados para igual contra menor		$\Phi \left(\frac{X_c - \mu_1}{\sigma / \sqrt{n}} \right)$ o bien $\tau_{n-1} \left(\frac{X_c - \mu_1}{S / \sqrt{n}} \right)$

Donde:

- α es el nivel de significación, es decir, la probabilidad de cometer el error tipo I (rechazar H_0 cuando en realidad es verdadera).
- \bar{X} es el promedio muestral.
- n es el tamaño de la muestra. Si la población no es normal, n debe ser mayor a 30. Si no se conoce el desvío poblacional pero $n \geq 50$, se puede usar el fractil z en vez del fractil t .
- X_c es el valor que está en el miembro derecho de la correspondiente inecuación. Para la prueba de igual contra distinto, X_{c2} es el miembro derecho de la inecuación superior, y X_{c1} es el miembro derecho de la inecuación inferior.
- Φ , τ_{n-1} son las funciones de distribución acumulada de la normal estándar y la t-Student con $n-1$ grados de libertad respectivamente. Se consiguen integrando o de las tablas.
- $z_{1-\alpha}$ y $t_{1-\alpha;n-1}$ son fractiles de la distribución normal estándar y la distribución t-Student respectivamente, que se obtienen de las tablas.
- μ_v es el valor "verdadero" de μ , que se usa para calcular la potencia.

Justificaciones

1) Igual contra mayor:

$$H_0: \mu = \mu_0$$

$$H_A: \mu > \mu_0$$

Como estudiamos en la primera sección del capítulo, α es la probabilidad de cometer el error tipo I (rechazar H_0 cuando en realidad es verdadera). En una prueba contra mayor, se determina un valor crítico (X_c) tal que si el estimador (en este caso, \bar{X}) es mayor que ese valor crítico, se rechaza H_0 (porque resulta más razonable la alternativa de que en realidad $\mu > \mu_0$).

Entonces la probabilidad de rechazar H_0 cuando en realidad es verdadera es:

$$P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = \alpha$$

Como rechazamos $H_0 \iff \bar{X} > X_c$ entonces queda:

$$P\left(\bar{X} > X_c / H_0 \text{ verdadera}\right) = \alpha$$

Que H_0 sea verdadera significa que $\mu = \mu_0$ con lo cual:

$$P\left(\bar{X} > X_c / \mu = \mu_0\right) = \alpha$$

La distribución de \bar{X} es $N\left(\mu ; \frac{\sigma}{\sqrt{n}}\right)$. Vemos que obviamente está en función del

μ desconocido. Ahora para poder resolver la probabilidad condicional que nos quedó, vamos a asumir que $\mu = \mu_0$, con lo cual la distribución de X condicionada

a ese valor de μ es $N\left(\mu_0 ; \frac{\sigma}{\sqrt{n}}\right)$, y entonces $P\left(X > X_c / \mu = \mu_0\right) = \alpha$ se reduce a $P(X > X_c) = \alpha$

Estandarizando queda:

$$1 - \Phi\left(\frac{X_c - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = \alpha$$

De donde:

$$\Phi\left(\frac{X_c - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha \Rightarrow \frac{X_c - \mu_0}{\frac{\sigma}{\sqrt{n}}} = z_{1-\alpha} \Rightarrow X_c = \mu_0 + \frac{z_{1-\alpha}\sigma}{\sqrt{n}}$$

Con lo cual hemos determinado el valor de X_c .

La potencia del test es $1 - \beta$, donde β es la probabilidad de cometer el error tipo II (no rechazar H_0 cuando en realidad es falsa). Calculemos β :

$$P\left(\text{no rechazar } H_0 / H_0 \text{ falsa}\right) = \beta$$

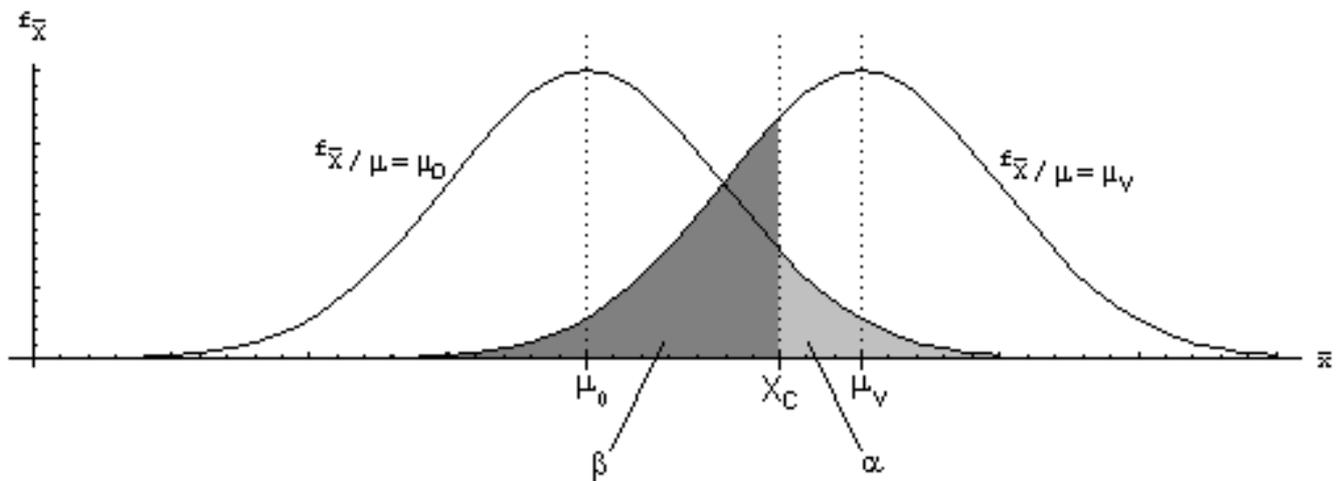
Como no rechazamos $H_0 \Leftrightarrow X < X_c$ entonces queda:

$$P\left(X < X_c / H_0 \text{ falsa}\right) = \beta$$

Que H_0 sea falsa significa que μ no vale μ_0 sino otro valor $\mu_v \neq \mu_0$. Es decir, en este caso, el error tipo II es: "no rechazar que ' $\mu = \mu_0$ ' cuando en realidad ' $\mu = \mu_v$ '".

Dicho valor μ_v puede ser elegido arbitrariamente y sirve para estudiar las propiedades del test que estamos llevando a cabo (normalmente deseamos que para valores μ_v razonablemente distintos de μ_0 , haya alta probabilidad de rechazar " $\mu = \mu_0$ ". Esto se puede estudiar observando los valores de β para distintos μ_v). Queda:

$$P\left(X < X_c / \mu = \mu_v\right) = \beta$$



Ahora asumiremos que la distribución de \bar{X} está condicionada a que $\mu = \mu_v$, con lo

cual usaremos $\bar{X} : N\left(\mu_v ; \frac{\sigma}{\sqrt{n}}\right)$ y entonces estandarizamos así:

$$\Phi\left(\frac{X_c - \mu_v}{\frac{\sigma}{\sqrt{n}}}\right) = \beta$$

La potencia es $1 - \beta$, con lo cual:

$$\text{Potencia} = 1 - \beta = 1 - \Phi\left(\frac{X_c - \mu_v}{\frac{\sigma}{\sqrt{n}}}\right) = \Phi\left(\frac{\mu_v - X_c}{\frac{\sigma}{\sqrt{n}}}\right)$$

Pero ahora volvamos al principio. Si no conociéramos el desvío poblacional, no podríamos utilizar la normal estándar. Llegaríamos hasta:

$$P\left(\bar{X} > X_c / \mu = \mu_0\right) = \alpha$$

$$Z = \frac{\bar{X} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

y después en vez de tomar

$$\Phi\left(\frac{X_c - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$$

tendríamos que tomar

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

y reemplazar $F_x(X_c)$ por

y reemplazar $F_x(X_c)$ por la integral

de la t-Student análoga a la que notamos como Φ para la normal estándar (la

notaremos τ), con lo cual $P(\bar{X} > X_c) = \alpha$ se convierte en:

$$1 - \tau\left(\frac{X_c - \mu_0}{\frac{S}{\sqrt{n}}}\right) = \alpha$$

y luego:

$$\tau\left(\frac{X_c - \mu_0}{S/\sqrt{n}}\right) = 1 - \alpha \Rightarrow \frac{X_c - \mu_0}{S/\sqrt{n}} = t_{1-\alpha; n-1} \Rightarrow X_c = \mu_0 + \frac{t_{1-\alpha; n-1} S}{\sqrt{n}}$$

Con lo cual el X crítico para cuando no conocemos el desvío poblacional nos quedó análogo al que usamos cuando sí lo conocemos.

La potencia, en vez de quedar:

$$\text{Potencia} = \Phi\left(\frac{\mu_v - X_c}{\sigma/\sqrt{n}}\right)$$

queda:

$$\text{Potencia} = \tau\left(\frac{\mu_v - X_c}{S/\sqrt{n}}\right)$$

Con lo cual hemos deducido las 4 fórmulas que dimos para las pruebas de igual contra mayor.

Si la población no es normal, valen las mismas aclaraciones que hicimos en el capítulo anterior para deducir los intervalos de confianza. Con $n \geq 30$, estos resultados se consideran válidos aunque la población no sea normal. Además, con $n \geq 50$, se puede usar el fractil z aunque no se conozca el desvío poblacional.

2) Igual contra menor:

$$H_0: \mu = \mu_0$$

$$H_A: \mu < \mu_0$$

Las deducciones son análogas y quedan a cargo del lector. La única diferencia es que, como se estudió en la primera sección de este capítulo, para una prueba por igual contra menor H_0 se rechaza si el estimador (en este caso \bar{X}) resulta menor que X_c (a diferencia de la prueba por igual contra mayor, en la cual H_0 se rechaza si el estimador resulta mayor que X_c).

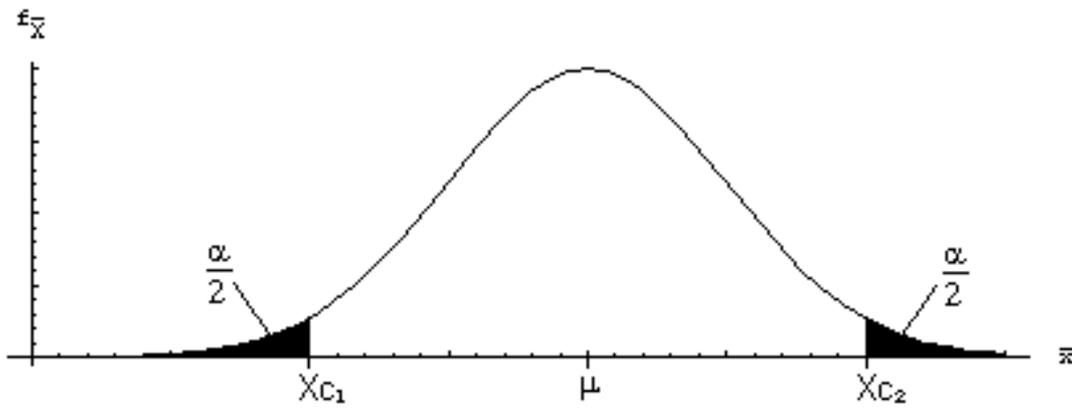
3) Igual contra distinto:

$$H_0: \mu = \mu_0$$

$$H_A: \mu \neq \mu_0$$

Las deducciones son también casi análogas, pero ahora α está repartida entre dos colas. Rechazamos H_0 si \bar{X} no cae dentro de ciertos límites X_{c1} y X_{c2} porque en ese caso \bar{X} queda tan lejos de μ_0 que resulta más razonable " $\mu \neq \mu_0$ " que " $\mu = \mu_0$ ".

$$P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = \alpha, \text{ quedando } \alpha \text{ repartida así:}$$



$$P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = P\left(X < X_{c1} / H_0 \text{ verdadera}\right) + P\left(X > X_{c2} / H_0 \text{ verdadera}\right) = \alpha$$

Por simplicidad, aprovecharemos la simetría de la normal para trabajar con un solo límite, y entonces:

$$P\left(X > X_{c2} / H_0 \text{ verdadera}\right) = \frac{\alpha}{2}$$

Luego, usando los mismos pasos intermedios que en la deducción de la fórmula para la prueba de igual contra mayor, llegamos a:

$$X_{c2} = \mu_0 + \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$$

que resulta igual a la mencionada fórmula con la diferencia de que ahora el fractil no es de $1-\alpha$ sino de $1-\alpha/2$.

Por último, encontraremos el otro límite por simetría:

$$X_{c1} = \mu_0 - \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}$$

La potencia del test es:

$$\text{Potencia} = 1 - \beta = 1 - P\left(\text{no rechazar } H_0 / H_0 \text{ falsa}\right) = 1 - P\left(X_{c1} < X < X_{c2} / H_0 \text{ falsa}\right) = 1 - P\left(X_{c1} < X < X_{c2} / \mu = \mu_v\right)$$

Asumiendo que $\mu = \mu_v$, queda:

$$1 - P\left(X_{c1} < X < X_{c2} / \mu = \mu_v\right) = 1 - P\left(X_{c1} < X < X_{c2}\right) = 1 - \left[\Phi\left(\frac{X_{c2} - \mu_v}{\sigma / \sqrt{n}}\right) - \Phi\left(\frac{X_{c1} - \mu_v}{\sigma / \sqrt{n}}\right) \right] = 1 - \Phi\left(\frac{X_{c2} - \mu_v}{\sigma / \sqrt{n}}\right) + \Phi\left(\frac{X_{c1} - \mu_v}{\sigma / \sqrt{n}}\right)$$

Las fórmulas para cuando no se conoce el desvío poblacional se deducen análogamente a como ya se estudió.

4) Igual contra igual

$$H_0: \mu = \mu_0$$

$$H_A: \mu = \mu_1$$

Si $\mu_0 < \mu_1$, los razonamientos y deducciones son todos análogos a los que se hacen para la prueba de igual contra mayor, y si $\mu_0 > \mu_1$, son análogos a los que se hacen

para la prueba de igual contra menor. Es decir, para $\mu_0 < \mu_1$ se determinará un X_c tal que se rechazaría H_0 si \bar{X} resultara mayor a X_c (porque resultaría más razonable pensar que el verdadero valor de μ es el valor más grande(μ_1)), y para $\mu_0 > \mu_1$ se determinará un X_c tal que se rechazaría H_0 si \bar{X} resultara menor a X_c (porque resultaría más razonable pensar que μ es el valor más chico(μ_0)).

La única diferencia consiste en que como estamos suponiendo por hipótesis que μ vale μ_0 o bien μ_1 , entonces el μ_v usado en la potencia es directamente μ_1 según corresponda, porque que H_0 sea falsa significa que μ no vale μ_0 , y entonces vale el único otro valor que suponemos posible: μ_1 .

Problemas típicos

1) Una empresa productora de leche afirma que su leche tiene 150 bacterias por mililitro. En un control de calidad, se extraen 9 muestras de 1ml, encontrándose las siguientes cantidades de bacterias:

140, 163, 152, 178, 145, 163, 149, 154, 153

La cantidad de bacterias en 1ml de leche es una variable aleatoria normal con desvío 10. Decida si a un nivel de significación del 1% se puede decir que la leche tiene más de 150 bacterias por mililitro. Indique la potencia del test si la verdadera media de la cantidad de bacterias por ml fuera 170.

Resolución:

$H_0: \mu = 150$

$H_A: \mu > 150$

Aunque podríamos usar directamente la regla de decisión dada (Rechazamos H_0 si $\bar{X} > \mu_0 + \frac{z_{1-\alpha} \sigma}{\sqrt{n}}$), y cuya deducción ya se exhibió, por esta vez haremos la deducción aplicada al ejemplo:

α es el nivel de significación, es decir, la probabilidad de cometer el error tipo I (rechazar H_0 cuando en realidad es verdadera). En este caso, $\alpha = 1\% = 0.01$.

En una prueba contra mayor para la media de una población normal, se determina un valor crítico (X_c) tal que si \bar{X} es mayor que ese valor crítico, se rechaza H_0 ($\mu = 150$) porque resulta más razonable la alternativa de que en realidad $\mu > 150$.

Entonces la probabilidad de rechazar H_0 cuando en realidad es verdadera es:

$$P\left(\frac{\text{rechazar } H_0}{H_0 \text{ verdadera}}\right) = \alpha = 0.01$$

Como rechazamos $H_0 \iff \bar{X} > X_c$, entonces queda:

$$P\left(\frac{\bar{X} > X_c}{H_0 \text{ verdadera}}\right) = 0.01$$

Que H_0 sea verdadera significa que $\mu = 150$ con lo cual:

$$P\left(\bar{X} > X_c \mid \mu = 150\right) = 0.01$$

La distribución de \bar{X} es $N\left(\mu ; \frac{10}{\sqrt{9}}\right)$. Vemos que obviamente está en función del μ desconocido. Ahora para poder resolver la probabilidad condicional que nos quedó, vamos a asumir que $\mu = 150$, con lo cual la distribución de \bar{X} condicionada

a ese valor de μ es $N\left(150 ; \frac{10}{\sqrt{9}}\right)$, y entonces $P\left(\bar{X} > X_c \mid \mu = 150\right) = 0.01$ se reduce a $P(\bar{X} > X_c) = 0.01$

Estandarizando queda:

$$1 - \Phi\left(\frac{X_c - 150}{\frac{10}{\sqrt{9}}}\right) = 0.01$$

De donde:

$$\Phi\left(\frac{X_c - 150}{\frac{10}{\sqrt{9}}}\right) = 1 - 0.01 \Rightarrow \frac{X_c - 150}{\frac{10}{\sqrt{9}}} = z_{0.99} \Rightarrow X_c = 150 + \frac{z_{0.99} \cdot 10}{\sqrt{9}} = 157.75$$

Con lo cual hemos determinado el valor de X_c .

Ahora calculamos: $\bar{X} = 155.22$

Vemos que $\bar{X} < X_c$, con lo cual no estamos en condiciones de rechazar H_0 . Es decir, no hay suficiente evidencia estadística como para decir que $\mu > 150$.

Ahora veamos cuál sería la potencia del test si el verdadero valor de μ fuera 170. Es decir, calcularemos la probabilidad de rechazar la hipótesis de que $\mu = 150$ cuando en efecto $\mu = 170$. Nuevamente, volveremos a deducir, aplicada a este ejemplo, la fórmula dada:

La potencia del test es $1 - \beta$, donde β es la probabilidad de cometer el error tipo II (no rechazar H_0 cuando en realidad es falsa). Calculemos β :

$$P\left(\text{no rechazar } H_0 \mid H_0 \text{ falsa}\right) = \beta$$

Como no rechazamos $H_0 \Leftrightarrow \bar{X} < X_c$ entonces queda:

$$P\left(\bar{X} < X_c \mid H_0 \text{ falsa}\right) = \beta$$

Que H_0 sea falsa significa que μ no vale 150 sino otro valor $\mu_v \neq \mu_0$. Usaremos $\mu_v = 170$. Es decir, en este caso, el error tipo II es: "no rechazar que ' $\mu = 150$ ' cuando en realidad ' $\mu = 170$ '". Nos interesa estudiar esto porque deseamos que si el verdadero valor de μ es razonablemente distinto de 150, haya alta probabilidad de rechazar la hipótesis " $\mu = 150$ ". Queda:

$$P\left(X < X_c / \mu = 170\right) = \beta$$

Ahora asumiremos que la distribución de X está condicionada a que $\mu = 170$, con

lo cual usaremos $X : N\left(170 ; \frac{10}{\sqrt{9}}\right)$ y entonces estandarizamos así:

$$\Phi\left(\frac{X_c - 170}{\frac{10}{\sqrt{9}}}\right) = \beta$$

X_C lo calculamos en el punto anterior y valía 157.75. La potencia es $1 - \beta$, con lo cual:

$$\text{Potencia} = 1 - \beta = 1 - \Phi\left(\frac{X_c - 170}{\frac{10}{\sqrt{9}}}\right) = \Phi\left(\frac{170 - X_c}{\frac{10}{\sqrt{9}}}\right) = \Phi(3,67) = 0,999$$

Vemos que el test es bueno, porque la probabilidad de rechazar que $\mu = 150$ en el caso de que el verdadero valor de μ fuera 170 es muy alta.

2) Una fábrica de gaseosas afirma que sus botellas contienen 2 litros de producto, pero se sospecha que en realidad contienen menos. La cantidad de líquido en las botellas es una variable aleatoria normal con desvío 0,1 litros. Un inspector toma una muestra de 10 botellas, y mide su contenido obteniendo un promedio por botella de 1,92 litros. Se pide:

- a) **Determine, a un nivel de significación del 5%, si hay evidencia para decir que las botellas contienen menos de 2 litros.**
- b) **¿Cómo se modifica lo hecho en a) si en realidad el desvío de 0,1 litros fue calculado a partir de la muestra?**

Resolución

Como ya hemos dado las fórmulas, su deducción y la aplicación de la deducción a un ejemplo, en adelante no volveremos a hacer las deducciones.

Estamos ante una prueba de igual contra menor:

$$H_0: \mu = 2$$

$$H_A: \mu < 2$$

a) σ es conocido, con lo cual rechazaremos H_0 si:

$$X < \mu_0 - \frac{z_{1-\alpha} \sigma}{\sqrt{n}}$$

$$\alpha = 0.05$$

$$\mu_0 - \frac{z_{1-\alpha} \sigma}{\sqrt{n}} = 2 - \frac{z_{0,95} 0,1}{\sqrt{10}} = 1.948$$

Como $\bar{X} = 1.92 < 1.948$, informamos que, a un nivel de significación del 5%, rechazamos que $\mu = 2$ contra la alternativa de que en realidad μ sea menor a 2.

b) σ es desconocido, y el valor que nos dieron en realidad es S. Rechazaremos H_0 si:

$$\bar{X} < \mu_0 - \frac{t_{1-\alpha; n-1} S}{\sqrt{n}}$$

$$\alpha = 0.05$$

$$\mu_0 - \frac{t_{1-\alpha; n-1} S}{\sqrt{n}} = 2 - \frac{1.8331 \cdot 0.1}{\sqrt{10}} = 1.942$$

Como $\bar{X} = 1.92 < 1.942$, informamos que, a un nivel de significación del 5%, rechazamos que $m = 2$ contra la alternativa de que en realidad μ sea menor a 2.

3) Se estudian 40 latas de arvejas, encontrándose que el peso promedio es de 220 g (y se sabe que el desvío poblacional es 25g).

a) ¿Es aceptable, a un nivel de significación del 1%, la hipótesis de que la media de la distribución es 200g?

b) Calcule la potencia del test si la verdadera media fuera 190.

Extraiga conclusiones y proponga qué se puede hacer.

Resolución

a) La población no es normal pero $n \geq 30$, con lo cual podemos usar las fórmulas que tenemos. Nos encontramos frente a una prueba de igual contra distinto, con desvío conocido:

$$H_0: \mu = 200$$

$$H_A: \mu \neq 200$$

Rechazaremos H_0 si \bar{X} queda fuera del intervalo:

$$\left(\mu_0 - \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} ; \mu_0 + \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} \right)$$

Usando los datos, el intervalo queda:

$$(189.8 ; 210.2)$$

Como $\bar{X} = 220$ ha quedado fuera del intervalo centrado en μ_0 , decidimos rechazar que $\mu = 200$ por resultar más razonable pensar que: $\mu \neq 200$.

$$\text{Potencia} = 1 - \Phi\left(\frac{X_{c2} - 190}{\frac{\sigma}{\sqrt{n}}}\right) + \Phi\left(\frac{X_{c1} - 190}{\frac{\sigma}{\sqrt{n}}}\right) = 0.48$$

b)

Este valor es demasiado bajo. Indica que, si la verdadera media fuera en realidad 190, habría solamente 48% de probabilidad de rechazar la hipótesis de que $\mu = 200$. Hay dos formas de mejorar este resultado:

- Aumentando el tamaño de la muestra:

Si tomamos $n = 80$ en vez de $n = 40$, el intervalo queda $(192.8 ; 207.2)$ y la potencia (si $\mu = 190$) es 0.84.

• Aumentando α (sacrificando un el nivel de significación para poder reducir β).

Si trabajamos con $\alpha = 0.05$ en vez de $\alpha = 0.01$, el intervalo queda $(192.3 ; 207.7)$ y la potencia (si $\mu = 190$) es 0.71.

Si hacemos las dos cosas propuestas, el intervalo queda $(194.5 ; 205.5)$ y la potencia (si $\mu = 190$) es 0.95.

4) Dos operarios de una fábrica son los únicos que saben cómo calibrar la máquina que embolsa maníes. Uno de ellos afirma que calibró la máquina para que produzca bolsas de 150g. El otro operario dice que el primero miente, y que en verdad la máquina fue calibrada para que produzca bolsas de 160g. Suponga el desvío poblacional conocido e igual a 20g.

a) Si se toma una muestra de 30 bolsas y el promedio de los pesos es 156 g., ¿Qué puede afirmar acerca de la declaración del primer operario, a un nivel de significación del 5%? Extraiga conclusiones acerca de este resultado.

b) Si la regla de decisión fuera decir que el primer operario miente si el promedio resulta mayor a 155g, ¿cuál es la probabilidad de cometer los errores tipo I y tipo II?.

Resolución

a) Estamos ante una prueba de igual contra igual:

$H_0: \mu = 150$

$H_A: \mu = 180$

No nos dicen que la población se puede ser supuesta normal, pero podemos proceder como si lo fuera porque $n \geq 30$. Como $150 < 180$, buscaremos un X_c tal que si el promedio resulta mayor que dicho valor crítico, decidiremos que el primer operario mintió. Esto es similar a lo que haríamos en una prueba de igual contra mayor, por lo cual el valor crítico será:

$$X_c = \mu_0 + \frac{z_{1-\alpha} \sigma}{\sqrt{n}} = 150 + \frac{1.645 \cdot 20}{\sqrt{30}} = 156.006$$

El promedio es 156, con lo cual no queda en la región de rechazo. Entonces no deberíamos rechazar H_0 . Sin embargo, queda muy cerca del borde, y no olvidemos que aunque podemos usar la aproximación normal por ser n grande, n tampoco era tan grande (cumplía con el 30 pedido, pero no era más grande que eso). En este caso, resultaría saludable aumentar un poco el n para mejorar la aproximación, y si lo hiciéramos, el X_c quedaría más cerca de 150 con lo cual si el promedio nos quedara similar al que obtuvimos al principio, quedaría probablemente en la región de rechazo, y deberíamos concluir que el primer operario mintió. La mejor opción en este caso es aumentar un poco el tamaño de la muestra y recalcular el promedio y el valor crítico y entonces decidir.

b) En este caso, en vez de darnos el α para que calculemos el X_c , nos están dando el X_c y nos preguntan el α .

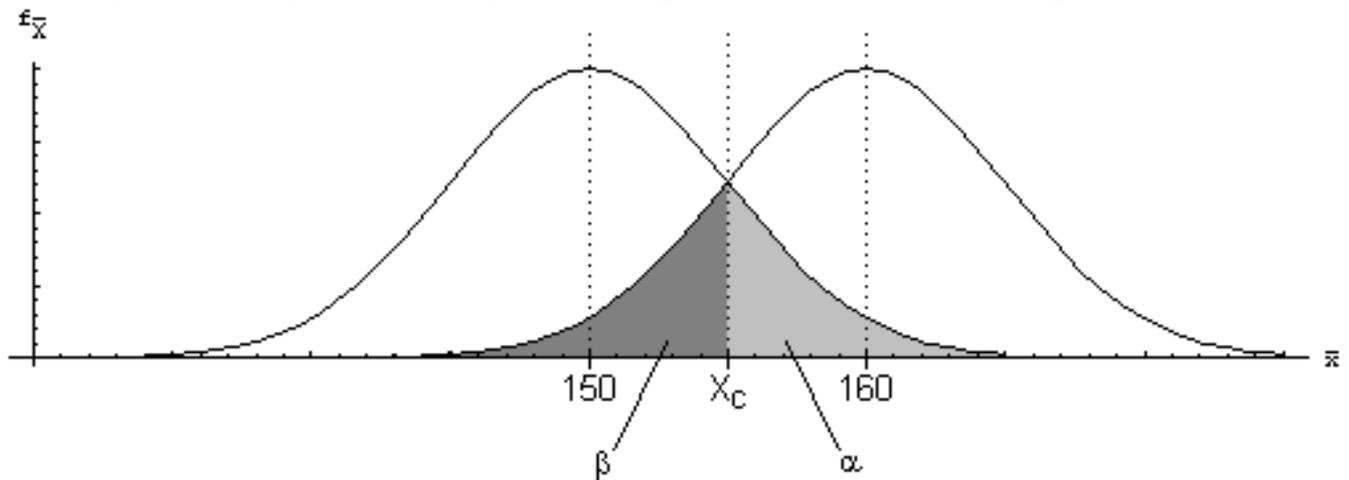
$$P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = \alpha$$

$$\alpha = P\left(X > X_c / \mu = 150\right) = 1 - \Phi\left(\frac{X_c - 150}{20/\sqrt{30}}\right) = 1 - \Phi\left(\frac{155 - 150}{20/\sqrt{30}}\right) = 0.085$$

Con respecto a la potencia, como la prueba es de igual contra igual, el suceso en el cual H_0 es falsa puede ser reescrito como que μ es igual al valor propuesto por H_A .

$$\beta = P\left(\text{no rechazar } H_0 / H_0 \text{ falsa}\right) = P\left(X < X_c / \mu = 160\right) = \Phi\left(\frac{X_c - 160}{20/\sqrt{30}}\right) = \Phi\left(\frac{155 - 160}{20/\sqrt{30}}\right) = 0.085$$

Hasta aquí lo que pedía el problema. Además podríamos observar que α y β quedaron iguales. Esto es porque el X_c está exactamente en el medio entre los dos valores propuestos por las dos hipótesis de igualdad. Veámoslo en un gráfico:



5) La longitud de los tornillos fabricados por una máquina es una variable aleatoria normal, cuyo desvío es 2 cm., y cuya media debería ser 12 cm. Su desvío es 2. Se sospecha que la máquina está descalibrada y produce tornillos más largos de lo debido. Se toma una muestra, y si el promedio resulta mayor a 13, se decide que la máquina está descalibrada.

a) ¿Cuántos tornillos habría que revisar para que la probabilidad de pensar que la máquina está bien calibrada cuando en realidad la media es 13.5 sea inferior a 0,01?

b) En ese caso, ¿cuál es la probabilidad de pensar que está descalibrada cuando en realidad no lo está?

Resolución

a) Prueba de igual contra mayor:

$$H_0: \mu = 12$$

$$H_A: \mu > 12$$

Luego la regla de decisión es rechazar H_0 si \bar{X} resulta mayor a un cierto X_c . El X_c ya está elegido y vale 13. Con ese X_c , se puede calcular la probabilidad de no rechazar H_0 en el caso de que μ en realidad fuera 14:

$$\beta = P\left(\text{no rechazar } H_0 / H_0 \text{ falsa}\right) = P\left(\bar{X} < 13 / H_0 \text{ falsa}\right) = P\left(\bar{X} < 13 / \mu = 13.5\right)$$

$$X: N\left(13.5 ; \frac{2}{\sqrt{n}}\right)$$

Suponiendo que $\mu = 13.5$, queda

$$P\left(\bar{X} < 13 / \mu = 13.5\right) = \Phi\left(\frac{13-13.5}{2/\sqrt{n}}\right) = \Phi\left(\frac{\sqrt{n}}{-4}\right)$$

Se pide que en estas circunstancias, $\beta < 0.01$, con lo cual:

$$\Phi\left(\frac{\sqrt{n}}{-4}\right) < 0.01 \Rightarrow \frac{\sqrt{n}}{-4} < z_{0.01} \Rightarrow n > (-4 z_{0.01})^2 \Rightarrow n > 86.6 \Rightarrow n = 87$$

Entonces, revisando 87 tornillos, la probabilidad de pensar que la máquina está calibrada cuando en realidad $\mu = 13.1$ es menor que 0.01.

$$b) P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = P\left(\bar{X} > X_c / H_0 \text{ verdadera}\right) = P\left(\bar{X} > 13 / \mu = 12\right)$$

$$X: N\left(14 ; \frac{2}{\sqrt{87}}\right)$$

Suponiendo que $\mu = 12$, queda

$$P\left(\bar{X} > 13 / \mu = 12\right) = 1 - \Phi\left(\frac{13-12}{2/\sqrt{87}}\right) = 1 - \Phi(4.66) \approx 0$$

Observamos que cuando n es grande, se puede tener al mismo tiempo α y β chicos.

6) Diseñe una prueba de hipótesis a dos colas para verificar si la media de la longitud de los tornillos fabricados por una máquina (variable aleatoria normal con desvío 2 cm.) es de 12 cm. La probabilidad de cometer el error tipo I debe ser 0.01, y la probabilidad de cometer el error tipo II en caso de que la máquina esté descalibrada en más de 1 cm debe ser como máximo 0.01.

Resolución

Las pruebas a dos colas son las de igual contra distinto.

$$H_0: \mu = 12$$

$$H_A: \mu \neq 12$$

Como se estudió anteriormente, se tomará un intervalo alrededor del 12, tal que la probabilidad de que \bar{X} quede fuera del intervalo (es decir, la probabilidad de rechazar que $\mu = 12$ aunque sea verdad) es α . Como se tienen dos colas, a cada

lado deberá haber una probabilidad de $\alpha/2$.

$$P\left(\bar{X} > X_{c2} / \mu = 12\right) = \frac{\alpha}{2}$$

Luego, haciendo las cuentas, y como figura en la tabla, el intervalo queda:

$$\left(12 - \frac{2z_{1-\frac{\alpha}{2}}}{\sqrt{n}} ; 12 + \frac{2z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right)$$

Como α debe valer 0.01, queda:

$$\left(12 - \frac{1.68}{\sqrt{n}} ; 12 + \frac{1.68}{\sqrt{n}}\right)$$

Piden que la probabilidad de pensar que la máquina está calibrada cuando en realidad la máquina esté descalibrada en más de 1 cm debe ser menor o igual a 0.01.

Escribimos:

$$P\left(X_{c1} < \bar{X} < X_{c2} / \mu > 13 \vee \mu < 11\right) < 0.01$$

Cuanto más cercano esté μ a 12, mayor es la probabilidad de cometer el error tipo II, porque μ no es 12 pero \bar{X} quedará más fácilmente dentro de la región de no-rechazo. Si μ puede ser mayor a 13 o menor a 11, y el máximo β se dará cuanto más cerca esté μ de 12, entonces β tendrá su máximo cuando μ valga 11 o 13. Al alejarse más μ de 12, será menos probable que \bar{X} quede fuera de la región crítica. Entonces si el n que encontramos satisface que $\beta < 0.01$ para $\mu = 11$ o $\mu = 13$, necesariamente también lo satisface para cualquier μ menor a 11 ó mayor que 13.

$$P\left(X_{c1} < \bar{X} < X_{c2} / \mu = 11\right) < 0.01$$

$$\Phi\left(\frac{X_{c2} - 11}{2/\sqrt{n}}\right) - \Phi\left(\frac{X_{c1} - 11}{2/\sqrt{n}}\right) = \Phi\left(\frac{12 + \frac{1.68}{\sqrt{n}} - 11}{2/\sqrt{n}}\right) - \Phi\left(\frac{12 - \frac{1.68}{\sqrt{n}} - 11}{2/\sqrt{n}}\right) = \Phi\left(\frac{\sqrt{n} + 1.68}{2}\right) - \Phi\left(\frac{\sqrt{n} - 1.68}{2}\right) < 0.01$$

Probamos con $n = 1$, $n = 2$, etc., y vemos que a partir de $n = 41$ se cumple la desigualdad.

El intervalo queda:

$$(11.74 ; 12.26)$$

Entonces nuestra prueba quedará enunciada así:

"Se toman 41 tornillos y se calcula el promedio de sus longitudes. Si dicho promedio está comprendido entre 11.74 y 12.26, consideraremos que la máquina está calibrada, y de lo contrario consideraremos que está descalibrada."

Pruebas de hipótesis para una proporción

Cuando la muestra sea grande usaremos fórmulas aproximadas, y cuando la muestra sea chica nos veremos obligados a usar fórmulas más exactas.

Cuando la muestra es grande podemos usar:

Hipótesis	Se rechaza H_0 si:	La potencia del test $(1-\beta)$ es:
$H_0: p = p_0$ $H_A: p > p_0$ "igual contra mayor"	$\hat{p} > p_0 + z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}}$	$\Phi \left(\frac{p_v - p_c}{\sqrt{\frac{p_v(1-p_v)}{n}}} \right)$
$H_0: p = p_0$ $H_A: p < p_0$ "igual contra menor"	$\hat{p} < p_0 - z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}}$	$\Phi \left(\frac{p_c - p_v}{\sqrt{\frac{p_v(1-p_v)}{n}}} \right)$
$H_0: p = p_0$ $H_A: p \neq p_0$ "igual contra distinto"	$\hat{p} > p_0 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$ o bien: $\hat{p} < p_0 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$	$1 - \Phi \left(\frac{p_{c2} - p_v}{\sqrt{\frac{p_v(1-p_v)}{n}}} \right) + \Phi \left(\frac{p_{c1} - p_v}{\sqrt{\frac{p_v(1-p_v)}{n}}} \right)$
$H_0: p = p_0$ $H_A: p = p_1$ "igual contra igual"	si $p_0 < p_1$, los criterios coinciden con los usados para igual contra mayor	$\Phi \left(\frac{p_1 - p_c}{\sqrt{\frac{p_1(1-p_1)}{n}}} \right)$
	si $p_0 > p_1$, los criterios coinciden con los usados para igual contra menor	$\Phi \left(\frac{p_c - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} \right)$

Donde:

- α es el nivel de significación, es decir, la probabilidad de cometer el error tipo I (rechazar H_0 cuando en realidad es verdadera).
- $\hat{p} = \frac{X}{n}$
- X es la cantidad de éxitos en la muestra.
- n es el tamaño de la muestra y debe ser por lo menos $n \geq 50$. Lo óptimo es $n \geq 100$. Si n fuera < 50 , se pueden usar las fórmulas exactas que se dan más abajo.

- p_c es el valor que está en el miembro derecho de la correspondiente inecuación. Para la prueba de igual contra distinto, p_{c2} es el miembro derecho de la inecuación superior, y p_{c1} es el miembro derecho de la inecuación inferior.
- Φ es la función de distribución acumulada de la normal estándar, que se consigue de las tablas.
- $Z_{1-\alpha}$ es el fractil de la distribución normal estándar que acumula a la izquierda un área α . Se obtiene de las tablas.
- p_v es el valor "verdadero" de p , que se usa para calcular la potencia.

Cuando la muestra es chica usaremos:

Hipótesis	Se rechaza H_0 si:
$H_0: p = p_0$ $H_A: p > p_0$ "igual contra mayor"	$X > X_c$ donde X_c es el menor entero tal que: $\sum_{i=0}^{X_c} \binom{n}{i} p_0^i (1-p_0)^{n-i} \geq 1-\alpha$
$H_0: p = p_0$ $H_A: p < p_0$ "igual contra menor"	$X < X_c$ donde X_c es el mayor entero tal que: $\sum_{i=0}^{X_c-1} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha$
$H_0: p = p_0$ $H_A: p \neq p_0$ "igual contra distinto"	$X < X_{c1} \quad o \quad X > X_{c2}$ donde X_{c1} es el mayor entero tal que: $\sum_{i=0}^{X_{c1}-1} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \frac{\alpha}{2}$ y X_{c2} es el menor entero tal que: $\sum_{i=0}^{X_{c2}} \binom{n}{i} p_0^i (1-p_0)^{n-i} \geq 1-\frac{\alpha}{2}$

Donde:

- α es el nivel de significación, es decir, la probabilidad de cometer el error tipo I (rechazar H_0 cuando en realidad es verdadera).
- $\hat{p} = \frac{X}{n}$
- X es la cantidad de éxitos en la muestra.
- n es el tamaño de la muestra. Si n es mayor a 50, pueden usar las fórmulas aproximadas dadas anteriormente.

Justificación

Las reglas de decisión dadas se deducen de manera análoga a las de la media de una población, como se estudió en la sección anterior. Con respecto a la aproximación del desvío usando \hat{p} en vez de p , deben hacerse las mismas consideraciones estudiadas en el capítulo anterior. A modo de ejemplo, se ofrece la deducción de las fórmulas para la prueba por igual contra mayor:

$$H_0: p = p_0$$

$$H_A: p > p_0$$

Se determina p_c , y luego si $\hat{p} > p_c$ se rechaza H_0 .

$$P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = \alpha$$

$$P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = P\left(\hat{p} > p_c / H_0 \text{ verdadera}\right) = P\left(\hat{p} > p_c / p = p_0\right)$$

Como se estudió en capítulos anteriores, $X: N\left(\mu p; \sqrt{np(1-p)}\right)$ si n es grande.

Luego como $\hat{p} = \frac{X}{n}$, entonces $\hat{p}: N\left(p; \sqrt{\frac{p(1-p)}{n}}\right)$. Asumiendo $p = p_0$ para simplificar la condicional, y estandarizando, obtenemos:

$$P\left(\hat{p} > p_c / p = p_0\right) = 1 - \Phi\left(\frac{p_c - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) = \alpha$$

Despejando p_c se obtiene:

$$p_c = p_0 + z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}}$$

Para calcular la potencia comenzaremos por plantear:

$$\beta = P\left(\text{no rechazar } H_0 / H_0 \text{ falsa}\right) = P\left(\hat{p} < p_c / H_0 \text{ falsa}\right) = P\left(\hat{p} < p_c / p = p_v\right)$$

$$\text{potencia} = 1 - \beta = 1 - P\left(\hat{p} < p_c / p = p_v\right)$$

Asumiendo $p = p_v$ para simplificar la condicional, y estandarizando, obtenemos:

$$\text{potencia} = 1 - P\left(\hat{p} < p_c\right) = 1 - \Phi\left(\frac{p_c - p_v}{\sqrt{\frac{p_v(1-p_v)}{n}}}\right) = \Phi\left(\frac{p_v - p_c}{\sqrt{\frac{p_v(1-p_v)}{n}}}\right)$$

Si n es chico, no podremos usar la aproximación normal para $X: \text{Bi}(n; p)$, y tampoco

tendrá sentido estimar p mediante $\hat{p} = \frac{X}{n}$ porque si n es chico la estimación no es confiable. En consecuencia trabajaremos directamente con X , la cantidad de éxitos de la muestra. Encontraremos un valor crítico X_c , tal que si X resulta mayor que ese valor, será más razonable suponer que $p > p_0$ que $p = p_0$. Planteamos:

$$P\left(\frac{X > X_c}{p = p_0}\right) = \alpha$$

Asumiendo que $p = p_0$, es decir, que $X:Bi(n;p_0)$, queda:

$$P(X > X_c) = \alpha$$

Es decir:

$$P(X \leq X_c) = 1 - \alpha$$

Luego, se puede obtener X_c de:

$$\sum_{i=0}^{X_c} \left[\binom{n}{i} p_0^i (1-p_0)^{n-i} \right] = 1 - \alpha$$

Por lo general, no existirá un valor entero de X_c que satisfaga la igualdad.

Adoptaremos el criterio de elegir el valor X_c que garantice que la probabilidad del error tipo 1 sea menor o igual que α . Entonces queda:

$$\sum_{i=0}^{X_c} \left[\binom{n}{i} p_0^i (1-p_0)^{n-i} \right] \geq 1 - \alpha$$

donde X_c es el menor entero que satisface la inecuación.

Problemas típicos

1) Un determinado partido declara que el 40% del padrón electoral tiene intención de votarlo. Se hace una encuesta de intención de voto a 500 personas, de las cuales 189 manifiestan su preferencia por el partido. ¿Hay evidencia estadística como para rechazar la declaración hecha por el partido, a un nivel de significación del 1%? ¿Cuál es la potencia del test si en realidad solamente el 30% del padrón tiene intención de votarlo? Saque conclusiones.

Resolución

Haremos una prueba de igual contra menor:

$$H_0: p = 0.4$$

$$H_A: p < 0.4$$

Rechazaremos H_0 si $\hat{p} < p_0 - z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}}$

$$p_0 - z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} = 0.4 - 2.33 \sqrt{\frac{0.4 \cdot 0.6}{500}} = 0.349$$

$$\hat{p} = \frac{189}{500} = 0.378$$

Vemos que no hay evidencia estadística como para rechazar la declaración del partido.

$$Potencia = \Phi \left(\frac{p_c - p_v}{\sqrt{\frac{p_v(1-p_v)}{n}}} \right) = \Phi \left(\frac{0.349 - 0.3}{\sqrt{\frac{0.3(1-0.3)}{500}}} \right) = 0.9916$$

Vemos que el test es bueno porque la potencia es cercana a 1.

2) Un determinado partido declara que el 40% del padrón electoral tiene intención de votarlo. Se hace una encuesta de intención de voto a 20 personas, de las cuales 8 manifiestan su preferencia por el partido. ¿Hay evidencia estadística como para rechazar la declaración hecha por el partido, a un nivel de significación del 10%? ¿Cuál es la potencia del test si en realidad solamente el 30% del padrón tiene intención de votarlo? Saque conclusiones.

Resolución

Haremos una prueba de igual contra menor:

$$H_0: p = 0.4$$

$$H_A: p < 0.4$$

Rechazaremos H_0 si $X < X_c$, donde X_c es el mayor entero tal que:

$$\sum_{i=0}^{X_c-1} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha$$

con $p_0 = 0.4$. Ahora queremos obtener X_c . Para ello, iremos sumando las probabilidades binomiales hasta superar la probabilidad $\alpha = 0.1$.

i	término de la sumatoria	probabilidad acumulada
0	0.00004	0.00004
1	0.00049	0.00052
2	0.00309	0.00361
3	0.01235	0.01596
4	0.03499	0.05095
5	0.07465	0.12560

El máximo X_c-1 que satisface la inecuación es 4. Luego, $X_c = 5$.

Entonces, rechazaremos H_0 si $X < 5$.

Como $X = 8$, concluimos que no hay evidencia estadística como para rechazar H_0 .

Ahora calcularemos la potencia para $p_v = 0.3$

$$\beta = P \left(\text{no rechazar } H_0 / H_0 \text{ falsa} \right) = P \left(\hat{p} > p_c / H_0 \text{ falsa} \right) = P \left(X \geq X_c / p = p_v \right)$$

$$potencia = 1 - \beta = 1 - \left(X \geq X_c / p = p_v \right)$$

Asumiendo $p = p_v$ para simplificar la condicional:

$$potencia = 1 - P(X \geq X_c) = P(X < X_c) = \sum_{i=0}^{X_c-1} \left[\binom{n}{i} p_0^i (1-p_0)^{n-i} \right] = 0.23751$$

Vemos que el test es malo, porque la potencia es un número más cercano al cero que al uno. Para mejorar el test, habría que encuestar a más personas o bien subir un poco el riesgo α . Como α ya es un riesgo alto (10%), lo más razonable resulta ser encuestar a más personas.

También deben considerarse problemas típicos los análogos a los dados en la sección anterior para las pruebas sobre la media de una población.

Pruebas de hipótesis para la varianza de una población

La población debe ser normal y/o la muestra debe ser grande.

Hipótesis	Se rechaza H_0 si:
$H_0: \sigma^2 = \sigma_0^2$ $H_A: \sigma^2 > \sigma_0^2$ "igual contra mayor"	$S^2 > \frac{\chi_{1-\alpha; n-1}^2 \sigma_0^2}{n-1}$
$H_0: \sigma^2 = \sigma_0^2$ $H_A: \sigma^2 < \sigma_0^2$ "igual contra menor"	$S^2 < \frac{\chi_{\alpha; n-1}^2 \sigma_0^2}{n-1}$
$H_0: \sigma^2 = \sigma_0^2$ $H_A: \sigma^2 \neq \sigma_0^2$ "igual contra distinto"	$S^2 > \frac{\chi_{1-\frac{\alpha}{2}; n-1}^2 \sigma_0^2}{n-1} \quad \text{o bien} \quad S^2 < \frac{\chi_{\frac{\alpha}{2}; n-1}^2 \sigma_0^2}{n-1}$

Donde:

- α es el nivel de significación, es decir, la probabilidad de cometer el error tipo I (rechazar H_0 cuando en realidad es verdadera).

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

- n es el tamaño de la muestra. Si la población no es normal, debe ser por lo menos $n \geq 30$.

- $\chi_{\frac{\alpha}{2}; n-1}^2$, $\chi_{1-\frac{\alpha}{2}; n-1}^2$ son fractiles de la distribución chi-cuadrado con $n-1$ grados de libertad, que se obtienen de las tablas.

Justificación

Las reglas de decisión dadas se deducen de manera análoga a las vistas en las secciones anteriores de este mismo capítulo, y respetando las mismas consideraciones que se hicieron en el capítulo anterior. A modo de ejemplo, se ofrece la deducción de las fórmulas para la prueba por igual contra mayor:

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_A: \sigma^2 > \sigma_0^2$$

Se determina σ_c^2 , y luego si $S^2 > \sigma_c^2$ se rechaza H_0 .

$$P\left(\text{rechazar } H_0 \mid H_0 \text{ verdadera}\right) = \alpha$$

$$P\left(\text{rechazar } H_0 / H_0 \text{ verdadera}\right) = P\left(S^2 > \sigma_c^2 / H_0 \text{ verdadera}\right) = P\left(S^2 > \frac{\sigma_c^2}{\sigma^2} = \sigma_0^2\right) = \alpha$$

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

Como se estudió en capítulos anteriores, si tomamos χ^2 , donde n es el tamaño de una muestra tomada de una población normal, S^2 es la varianza muestral y σ^2 es la varianza poblacional, entonces χ^2 tiene una distribución chi-cuadrada con grados de libertad $\nu = n - 1$.

Para la distribución chi-cuadrado podemos escribir:

$$P(\chi^2 > \chi_{1-\alpha; n-1}^2) = \alpha$$

Usando el reemplazo que tomamos:

$$P\left(\frac{(n-1)S^2}{\sigma^2} > \chi_{1-\alpha; n-1}^2\right) = \alpha$$

Con lo cual:

$$P\left(S^2 > \frac{\chi_{1-\alpha; n-1}^2 \sigma^2}{n-1}\right) = \alpha$$

Si lo condicionamos a $\sigma^2 = \sigma_0^2$, queda:

$$P\left(S^2 > \frac{\chi_{1-\alpha; n-1}^2 \sigma_0^2}{n-1}\right) = \alpha$$

De donde observamos que:

$$\sigma_c^2 = \frac{\chi_{1-\alpha; n-1}^2 \sigma_0^2}{n-1}$$

Luego rechazamos H_0 si $S^2 > \frac{\chi_{1-\alpha; n-1}^2 \sigma_0^2}{n-1}$

Problemas típicos

1) De una población normal se extrae una muestra de tamaño 10 que arroja los siguientes valores:

25.3, 26.9, 28.7, 24.8, 30.2, 21.3, 22.4, 23.5, 22, 30.1

Determine si a un nivel de significación del 5% se puede afirmar que la varianza de la población es igual a 11, contra la alternativa de que sea mayor.

Resolución

Haremos una prueba de igual contra mayor:

$$H_0: \sigma^2 = 11$$

$$H_A: \sigma^2 > 11$$

$$S^2 > \frac{\chi_{1-\alpha; n-1}^2 \sigma_0^2}{n-1}$$

Rechazaremos H_0 si

Calculamos:

$$\bar{X} = 25.52$$

$$S^2 = \frac{\sum_{i=1}^{n-1} (x_i - \bar{X})^2}{n-1} = 11.05$$

$$\frac{\chi_{1-\alpha; n-1}^2 \sigma_0^2}{n-1} = \frac{16.919 \cdot 11}{9} = 20.68$$

Vemos que como 11.05 es bastante menor que 20.68, no solamente no hay evidencia como para afirmar que σ^2 sea mayor a 11, sino que además puede ser muy razonable afirmar que vale 11.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 12 de julio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 22 de julio de 2004_

Pruebas de hipótesis para las medias de dos poblaciones

Las poblaciones deben ser normales y/o las muestras deben ser grandes.

La hipótesis nula es $H_0: \mu_1 - \mu_2 = d$

La siguiente es una tabla de los criterios para rechazar H_0 frente a distintas alternativas. Las fórmulas a utilizar dependen de si se conocen o no se conocen las varianzas poblacionales σ_1^2 y σ_2^2 , y en el caso de que no se las conozca, si se las considera iguales o distintas.

σ_1^2 y σ_2^2	se calcula	hipótesis alternativa H_A		
		$\mu_1 - \mu_2 > d$	$\mu_1 - \mu_2 < d$	$\mu_1 - \mu_2 \neq d$
conocidas	$z = \frac{\bar{X}_1 - \bar{X}_2 - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$Z > Z_{1-\alpha}$	$Z < Z_\alpha$	$Z > Z_{1-\alpha/2}$ ó $Z < Z_{\alpha/2}$
desconocidas, pero se supone que $\sigma_1^2 = \sigma_2^2$	$t = \frac{\bar{X}_1 - \bar{X}_2 - d}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1 + n_2 - 2}}}$ $v = n_1 + n_2 - 2$	$t > t_{1-\alpha;v}$	$t < t_{\alpha;v}$	$t > t_{1-\alpha/2;v}$ ó $t < t_{\alpha/2;v}$
desconocidas, pero se supone que $\sigma_1^2 \neq \sigma_2^2$	$t = \frac{\bar{X}_1 - \bar{X}_2 - d}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ $v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$	$t > t_{1-\alpha;v}$	$t < t_{\alpha;v}$	$t > t_{1-\alpha/2;v}$ ó $t < t_{\alpha/2;v}$

Donde:

- α es el nivel de significación, es decir, la probabilidad de cometer el error tipo I (rechazar H_0 cuando en realidad es verdadera).
- n_1 y n_2 son los tamaños de las muestras de las poblaciones 1 y 2 respectivamente.

- \bar{X}_1 y \bar{X}_2 son los promedios de las muestras.
- S_1^2 y S_2^2 son las varianzas muestrales.
- z_α y $t_{\alpha, \nu}$ son fractiles de la distribución normal estándar y la distribución t-Student respectivamente, que se obtienen de las tablas.

Problemas típicos

1) Se cree que las mujeres argentinas son en promedio más altas que las uruguayas. Para eso se toma una muestra de 100 mujeres argentinas y 80 uruguayas, obteniéndose los promedios de alturas 1,65 y 1,60 respectivamente. Se sabe además que el desvío poblacional de la altura de las argentinas es 0,20 y el desvío poblacional de la altura de las uruguayas es 0,15. Decida a un nivel de significación del 5%.

Resolución

Las hipótesis que tenemos son:

$$H_0: \mu_A = \mu_U$$

$$H_A: \mu_A > \mu_U$$

Para respetar el formato de las fórmulas dadas, lo escribiremos así:

$$H_0: \mu_A - \mu_U = 0$$

$$H_A: \mu_A - \mu_U > 0$$

Como las varianzas poblacionales son conocidas, estamos frente al primer caso de la tabla. Calculamos z:

$$z = \frac{\bar{X}_A - \bar{X}_U - d}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_U^2}{n_U}}} = \frac{1.65 - 1.60 - 0}{\sqrt{\frac{0.2^2}{100} + \frac{0.15^2}{80}}} = 1.916$$

Por su parte, el fractil es $z_{1-\alpha} = z_{0.95} = 1.645$

Luego, como $z > z_{1-\alpha}$, rechazamos H_0 porque la información que tenemos indica que H_A resulta más razonable.

2) Se cree que las mujeres argentinas son en promedio más altas que las uruguayas. Para eso se toma una muestra de 100 mujeres argentinas y 80 uruguayas, obteniéndose los promedios de alturas 1,65 y 1,60 respectivamente. Se estiman las varianzas y se obtienen $S_A^2 = 0,05$ y $S_U^2 = 0,02$. Decida a un nivel de significación del 5%.

Resolución

Las hipótesis que tenemos, adecuadas a los formatos que usamos, son:

$$H_0: \mu_A - \mu_U = 0$$

$$H_A: \mu_A - \mu_U > 0$$

Como las varianzas poblacionales son desconocidas, antes de poder hacer el test tendremos que especificar si son iguales o no. Pero, lógicamente, si no las conocemos, no podemos saberlo.

¿Cómo se procede entonces? Sabemos que las varianzas muestrales son estimadores de las poblacionales. Si las varianzas muestrales son muy distintas, entonces probablemente las varianzas poblacionales sean distintas, y por lo tanto procederemos suponiendo que las varianzas poblacionales desconocidas son distintas. Si, en cambio, las varianzas muestrales arrojan valores muy parecidos, podría ser razonable suponer que las varianzas poblacionales son iguales.

¿Cuándo son lo suficientemente parecidas o distintas las muestrales? En la sección 10.7 se explica cómo hacer un test para determinarlo. Por ahora, en este ejemplo haremos la resolución de los dos casos.

a) Suponiendo que $\sigma_A^2 = \sigma_U^2$

Calculamos t:

$$t = \frac{\bar{X}_A - \bar{X}_U - d}{\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_U}\right) \frac{S_A^2(n_A - 1) + S_U^2(n_U - 1)}{n_A + n_U - 2}}} = \frac{1.65 - 1.60 - 0}{\sqrt{\left(\frac{1}{100} + \frac{1}{80}\right) \frac{0.05(100 - 1) + 0.02(80 - 1)}{100 + 80 - 2}}} = 1.740$$

Calculamos v:

$$v = n_A + n_U - 2 = 178$$

El fractil a considerar es $t_{1-\alpha;v}$. Como $v > 50$, lo aproximaremos con el fractil normal, es decir, $t_{1-\alpha;v} = t_{0.95;178} \cong z_{0.95} = 1.645$

Luego, como $t > t_{1-\alpha;v}$, rechazamos H_0 porque la información que tenemos indica que H_A resulta más razonable.

b) Suponiendo que $\sigma_A^2 \neq \sigma_U^2$

Calculamos t:

$$t = \frac{\bar{X}_A - \bar{X}_U - d}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_U^2}{n_U}}} = \frac{1.65 - 1.60 - 0}{\sqrt{\frac{0.05}{100} + \frac{0.02}{80}}} = 1.826$$

Calculamos v:

$$v = \frac{\left(\frac{S_A^2}{n_A} + \frac{S_U^2}{n_U}\right)^2}{\frac{\left(\frac{S_A^2}{n_A}\right)^2}{n_A - 1} + \frac{\left(\frac{S_U^2}{n_U}\right)^2}{n_U - 1}} = \frac{\left(\frac{0.05}{100} + \frac{0.02}{80}\right)^2}{\frac{\left(\frac{0.05}{100}\right)^2}{100 - 1} + \frac{\left(\frac{0.02}{80}\right)^2}{80 - 1}} = 169.6$$

El fractil a considerar es $t_{1-\alpha;v}$. Como $v > 50$, lo aproximaremos con el fractil normal, es decir, $t_{1-\alpha;v} = t_{0.95;178} \cong z_{0.95} = 1.645$

Luego, como $t > t_{1-\alpha;v}$, rechazamos H_0 porque la información que tenemos indica que H_A resulta más razonable.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 22 de julio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 22 de julio de 2004_

Pruebas de hipótesis para dos proporciones

Las muestras deben ser grandes, en lo posible mayores a 100.

Primero se calcula el estadístico z:

$$z = \frac{\hat{p}_1 - \hat{p}_2 - d}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

Y luego se usan los siguientes criterios:

Hipótesis	se rechaza H_0 si
$H_0: p_1 - p_2 = d$ $H_A: p_1 - p_2 > d$ "contra mayor"	$Z > Z_{1-\alpha}$
$H_0: p_1 - p_2 = d$ $H_A: p_1 - p_2 < d$ "contra menor"	$Z < Z_\alpha$
$H_0: p_1 - p_2 = d$ $H_A: p_1 - p_2 \neq d$ "contra distinto"	$Z > Z_{1-\alpha/2}$ ó $Z < Z_{\alpha/2}$

Donde:

- α es el nivel de significación, es decir, la probabilidad de cometer el error tipo I (rechazar H_0 cuando en realidad es verdadera).
- n_1 y n_2 son los tamaños de las muestras de las poblaciones 1 y 2 respectivamente.

- $\hat{p}_1 = \frac{X_1}{n_1}$; $\hat{p}_2 = \frac{X_2}{n_2}$

- X_1 y X_2 son la cantidad de éxitos en las muestras 1 y 2 respectivamente.
- Los z_α son fractiles de la distribución normal estándar, que se obtienen de las tablas.

Problemas típicos

1) Se cree que en la provincia de Santa Fe hay más gente con ojos claros que en la provincia de Entre Ríos. Para eso se toma una muestra de 300 personas de Sante Fe y 200 de Entre Ríos, obteniéndose 63 y 30 personas con ojos claros respectivamente. Decida a un nivel de significación del 5%.

Resolución

Las hipótesis que tenemos son:

$$H_0: p_S = p_E$$

$$H_A: p_S > p_E$$

Para respetar el formato de las fórmulas dadas, lo escribiremos así:

$$H_0: p_S - p_E = 0$$

$$H_A: p_S - p_E > 0$$

Calculamos las p muestrales:

$$\hat{p}_S = \frac{X_S}{n_S} = \frac{63}{300} = 0.21 \quad ; \quad \hat{p}_E = \frac{X_E}{n_E} = \frac{30}{200} = 0.15$$

Calculamos z:

$$z = \frac{\hat{p}_S - \hat{p}_E - d}{\sqrt{\frac{\hat{p}_S(1-\hat{p}_S)}{n_S} + \frac{\hat{p}_E(1-\hat{p}_E)}{n_E}}} = \frac{0.21 - 0.15 - 0}{\sqrt{\frac{0.21(1-0.21)}{300} + \frac{0.15(1-0.15)}{200}}} = 1.739$$

Por su parte, el fractil es $z_{1-\alpha} = z_{0.95} = 1.645$

Luego, como $z > z_{1-\alpha}$, rechazamos H_0 porque la información que tenemos indica que H_A resulta más razonable.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 22 de julio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 22 de julio de 2004_

Pruebas de hipótesis para dos varianzas

Estos tests sirven para determinar si las varianzas de dos poblaciones son iguales, mayores, menores o distintas.

Las poblaciones deben ser normales, o las muestras muy grandes.

Primero se calcula el estadístico F:

$$F = \frac{S_1^2}{S_2^2}$$

y los grados de libertad:

$$V_1 = n_1 - 1$$

$$V_2 = n_2 - 1$$

Y luego se usan los siguientes criterios:

Hipótesis	se rechaza H_0 si
$H_0 : \sigma_1^2 = \sigma_2^2$ $H_A : \sigma_1^2 > \sigma_2^2$ "contra mayor"	$F > f_{1-\alpha;v_1;v_2}$
$H_0 : \sigma_1^2 = \sigma_2^2$ $H_A : \sigma_1^2 < \sigma_2^2$ "contra menor"	$F < f_{\alpha;v_1;v_2}$
$H_0 : \sigma_1^2 = \sigma_2^2$ $H_A : \sigma_1^2 \neq \sigma_2^2$ "contra distinto"	$F > f_{1-\alpha/2;v_1;v_2}$ ó $F < f_{\alpha/2;v_1;v_2}$

Donde:

- α es el nivel de significación, es decir, la probabilidad de cometer el error tipo I (rechazar H_0 cuando en realidad es verdadera).
- n_1 y n_2 son los tamaños de las muestras de las poblaciones 1 y 2 respectivamente.
- S_1^2 y S_2^2 son las varianzas muestrales.
- $f_{\alpha;v_1;v_2}$ es el fractil de la distribución F con grados de libertad V_1 y V_2 con un área α acumulada a izquierda.

- Para conocer la distribución F y sus propiedades, ver la sección 6 del capítulo 7.
- La tabla de fractiles de la distribución F se encuentra en la sección 5 del apéndice D.

Problemas típicos

1) Se requiere saber si las varianzas de dos poblaciones son iguales o no. Para ello se toman dos muestras 1 y 2, de tamaños 100 y 80 respectivamente. Se calculan las varianzas muestrales y se obtienen 0,05 y 0,02 respectivamente. Decida a un nivel de significación del 1%.

Resolución

Las hipótesis que tenemos son:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_A : \sigma_1^2 \neq \sigma_2^2$$

Calculamos F y los grados de libertad:

$$F = \frac{S_1^2}{S_2^2} = \frac{0.05}{0.02} = 2.5$$

$$v_1 = n_1 - 1 = 99$$

$$v_2 = n_2 - 1 = 79$$

Por su parte, los fractiles son:

$$f_{1-\alpha/2; v_1; v_2} = f_{0.995; 99; 79} \cong 1.75$$

$$f_{\alpha/2; v_1; v_2} = f_{0.005; 99; 79} = 1 / f_{0.995; 79; 99} \cong 1 / 1.73 = 0.58$$

Como $F > f_{1-\alpha/2; v_1; v_2}$, rechazamos H_0 y concluimos que las varianzas poblacionales son distintas.

2) Se toma una muestra de 100 piezas producidas por una máquina y se obtiene que la varianza muestral es 15,4. Se necesita que la variabilidad de los pesos sea menor, por lo cual se somete a la máquina a un proceso de ajuste. Luego de ello, se toma otra muestra de 150 piezas producidas con la máquina ajustada, encontrándose que la varianza muestral es 11,2. ¿Puede afirmar que el ajuste ha sido efectivo? Es decir, ¿se ha realmente logrado reducir la varianza? Decida a un nivel de significación del 5%.

Resolución

Las hipótesis que tenemos son:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_A : \sigma_1^2 > \sigma_2^2$$

Calculamos F y los grados de libertad:

$$F = \frac{S_1^2}{S_2^2} = \frac{15.4}{11.2} = 1.375$$

$$v_1 = n_1 - 1 = 99$$

$$v_2 = n_2 - 1 = 149$$

Por su parte, el fractil es:

$$f_{1-\alpha;v_1;v_2} = f_{0.95;99;149} \cong 1.35$$

Como $F > f_{1-\alpha;v_1;v_2}$, rechazamos H_0 y concluimos que el ajuste ha sido efectivo.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 22 de julio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 12 de julio de 2004_

Prueba de bondad de ajuste

Esta prueba es un poco diferente de las demás pruebas estudiadas en este capítulo, porque las hipótesis no son sólo sobre los parámetros de determinadas distribuciones sino también sobre distribuciones en sí.

La prueba de bondad de ajuste sirve para determinar si es razonable pensar que determinados datos recogidos provienen de una determinada distribución específica.

Ejemplo con distribución discreta

Durante 30 días se ha registrado la cantidad de veces por día que un determinado servidor ha caído en downtime. Se obtuvieron los siguientes resultados:

0, 2, 4, 2, 3, 0, 1, 1, 2, 1, 3, 1, 2, 0, 2, 1, 1, 3, 2, 3, 2, 6, 0, 4, 1, 4, 2, 1, 3, 2.

¿Puede afirmarse, a un nivel de significación del 5%, que la cantidad de fallas por día sigue una distribución Poisson con $\mu = 2$?

Ejemplo con distribución continua

Se ha registrado la duración en horas de 25 componentes electrónicos. Se obtuvieron los siguientes resultados:

6.88, 17.68, 9.74, 20.05, 16.43, 19.99, 10.84, 3.54, 1.37, 5.87, 3.26, 4.20, 35.01, 8.45, 7.28, 4.32, 3.30, 4.20, 10.70, 7.93, 21.18, 19.33, 8.26, 5.03, 1.46

¿Puede afirmarse, a un nivel de significación del 5%, que la duración de los componentes tiene una distribución exponencial negativa con $\lambda = 0.1$?

Método

1) Debe conocerse la distribución que se desea probar, aunque no necesariamente el valor de sus parámetros. (Ejemplo: podemos suponer una distribución Poisson, aunque no necesariamente saber qué valor proponer para el parámetro μ).

2) Si no se sabe qué valores proponer para uno o más parámetros, estimarlos a partir de la muestra. Se recomienda usar el estimador habitual de ese parámetro, o bien estimar mediante máxima verosimilitud.

3) Plantear las hipótesis:

H_0 : Los datos recogidos provienen de una distribución xxxx con parámetro(s) xxxx.

H_A : Los datos recogidos no provienen de tal distribución.

4) Si la distribución a probar es discreta:

Hacer una tabla con la cantidad de veces que aparece en la muestra cada uno de los valores posibles de la distribución.

Si la distribución a probar es continua:

Dividir los valores posibles de la distribución en intervalos, y hacer una tabla con la frecuencia de cada uno de los intervalos (es decir, la cantidad de elementos de la muestra que caen en cada uno de los intervalos). En principio, hacer los intervalos con longitudes pequeñas.

En cualquiera de los dos casos, la tabla debe cumplir con lo siguiente:

- Si hay filas con frecuencia cero, se deben agrupar con el anterior o el siguiente, de modo tal que no queden filas con frecuencia cero.
- Si hay filas con frecuencias muy pequeñas (en relación con las demás) puede ser conveniente agruparlas de modo tal que no queden filas con frecuencias muy pequeñas.

5) Calcular para cada fila de la tabla la probabilidad de que una variable aleatoria distribuida según la distribución que se desea probar asuma el valor o los valores agrupados en dicha fila.

6) Calcular el estadístico:

$$X^2 = \sum_{i=1}^k \frac{(n p_i - x_i)^2}{n p_i}$$

donde:

- k es la cantidad de filas de la tabla
- x_i es la frecuencia de la fila
- p_i es la probabilidad de la fila
- n es el tamaño de la muestra

7) Rechazar H_0 si $X^2 > \chi^2_{1-\alpha; k-c}$

donde:

- α es el nivel de confianza
- k es la cantidad de filas de la tabla
- c es $1 +$ la cantidad de parámetros que fueron estimados en base a la muestra para poder proponer la distribución.

Resolución del ejemplo con distribución discreta

1) Vamos a ensayar si los datos recogidos vienen de una distribución Poisson con $\mu = 2$.

2) No necesitamos estimar ningún parámetro. El único parámetro de la distribución de Poisson es μ , y ya sabemos qué valor vamos a proponer para él.

3) Planteamos:

H_0 : Los datos recogidos provienen de una distribución Poisson con parámetro $\mu = 2$

H_A : Los datos recogidos no provienen de tal distribución.

4) Armamos la tabla de frecuencias:

Caídas	Frecuencia
0	4
1	8
2	9
3	5
4	3
5	0
6	1
7	0
8	0
...	0

Hay filas con frecuencia cero. Si agrupamos los ceros del 7 en adelante con el 6, la tabla queda así:

Caídas	Frecuencia
0	4
1	8
2	9
3	5
4	3
5	0
6 ó más	1

Sigue quedando una fila con cero. Agrupemos el "5" con el "6 ó más":

Caídas	Frecuencia
0	4
1	8
2	9
3	5
4	3
5 ó más	1

- Ya no queda ninguna fila con frecuencia cero.
- Observamos que ninguna fila ha quedado con frecuencia extremadamente pequeña con respecto a las demás.

Entonces hemos construido una tabla que cumple con las dos condiciones.

5) Calculamos la probabilidad de los valores agrupados en cada fila de la tabla:

$$P(X = 0) = \frac{e^{-2} 2^0}{0!} = 0,13534$$

$$P(X = 1) = \frac{e^{-2} 2^1}{1!} = 0,27067$$

$$P(X = 2) = \frac{e^{-2} 2^2}{2!} = 0,27067$$

$$P(X = 3) = \frac{e^{-2} 2^3}{3!} = 0,18045$$

$$P(X = 4) = \frac{e^{-2} 2^4}{4!} = 0,09022$$

$$P(X \geq 5) = 1 - P(X < 5) = 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4) = 0,05265$$

Le agregamos estas probabilidades a la tabla:

Caídas	Frecuencia	Probabilidad
0	4	0,13534
1	8	0,27067
2	9	0,27067
3	5	0,18045
4	3	0,09022
5 ó más	1	0,05265

$$X^2 = \sum_{i=1}^k \frac{(n p_i - x_i)^2}{n p_i} = 0,37397$$

6) Calculamos:

7) $\alpha = 0,05$; $k = 6$; $c = 1$, porque no se estimó ningún parámetro.

$$\chi_{0,95;5}^2 = 11,071$$

Buscamos en la tabla:

$$X^2 < \chi_{1-\alpha; k-c}^2$$

Como $X^2 < \chi_{1-\alpha; k-c}^2$, no rechazamos H_0 , y por lo tanto con un nivel de significación del 5% decimos que los datos recogidos efectivamente provienen de una distribución de Poisson con $\mu = 2$.

Resolución del ejemplo con distribución continua

1) Vamos a ensayar si los datos recogidos provienen de una distribución exponencial negativa con $\lambda = 0.1$

2) No necesitamos estimar ningún parámetro. El único parámetro de la distribución exponencial negativa es λ , y ya sabemos qué valor vamos a proponer para él.

3) Planteamos:

H_0 : Los datos recogidos provienen de una distribución Exponencial negativa con parámetro $\lambda = 0.1$

H_A : Los datos recogidos no provienen de tal distribución.

4) Elegimos intervalos 0-1.99, 2-3.99, 4-5.99, etc. y la tabla queda:

Duración	Frecuencia	Duración	Frecuencia
----------	------------	----------	------------

0 - 1.99	2
2 - 3.99	3
4 - 5.99	5
6 - 7.99	3
8 - 9.99	3
10 - 11.99	2
12 - 13.99	0
14 - 15.99	0
16 - 17.99	2

18 - 19.99	2
20 - 21.99	2
22 - 23.99	0
24 - 25.99	0
26 - 27.99	0
28 - 29.99	0
30 - 31.99	0
32 - 33.99	0
34 - 35.99	1

Quedaron algunos intervalos con frecuencia cero. Para solucionarlo, agruparemos algunos intervalos, y la tabla nos queda:

Duración	Frecuencia
0 - 1.99	2
2 - 3.99	3
4 - 5.99	5
6 - 7.99	3
8 - 9.99	3
10 - 13.99	2
14 - 17.99	2
18 - 19.99	2
20 - 21.99	2
22 ó más	1

- Ya no queda ninguna fila con frecuencia cero.
- Observamos que ninguna fila ha quedado con frecuencia extremadamente pequeña con respecto a las demás.

Entonces hemos construido una tabla que cumple con las condiciones pedidas.

5) Calculamos la probabilidad de que una variable distribuida exponencialmente con $\lambda = 0.1$ caiga en los intervalos de cada fila, y por comodidad agregamos dichos valores a la tabla:

Duración	Frecuencia	Probabilidad
0 - 1.99	2	0.18127
2 - 3.99	3	0.14841
4 - 5.99	5	0.12151
6 - 7.99	3	0.09948
8 - 9.99	3	0.08145
10 - 13.99	2	0.12128
14 - 17.99	2	0.08130
18 - 19.99	2	0.02996
20 - 21.99	2	0.02453
22 ó más	1	0.11080

$$X^2 = \sum_{i=1}^k \frac{(n p_i - x_i)^2}{n p_i} = 10,087$$

6) Calculamos:

7) $\alpha = 0,05$; $k = 10$; $c = 1$, porque no se estimó ningún parámetro.

$$\chi_{0,95;9}^2 = 16,919$$

Buscamos en la tabla:

Como $X^2 < \chi_{1-\alpha;k-c}^2$, no rechazamos H_0 , y por lo tanto con un nivel de significación del 5% decimos que los datos recogidos efectivamente provienen de una distribución exponencial negativa con $\lambda = 2$.

Problemas típicos

Además de los dos ejemplos ya resueltos, deben considerarse problemas típicos aquellos en los cuales hay que estimar el valor de los parámetros antes de poder hacer la prueba. A continuación, un ejemplo de ello:

1) En una determinada ciudad, las precipitaciones para el mes de abril de los últimos 40 años han sido:

12.62, 6.54, 7.00, 5.24, 9.98, 10.23, 11.79, 6.13, 6.82, 10.22, 6.58, 6.31, 10.88, 7.82, 6.61, 4.22, 6.72, 10.56, 9.66, 5.16, 7.14, 14.78, 10.46, 0.48, 8.94, 3.96, 1.84, 11.83, 10.07, 9.39, 1.78, 7.35, 5.81, 8.11, 9.71, 9.39, 7.73, 7.81, 9.20, 2.13

Determine, a un nivel de significación del 5%, si las precipitaciones siguen una distribución normal.

Resolución

Vamos a probar si las precipitaciones siguen una distribución normal. Para hacer una prueba de bondad de ajuste necesitamos probar una distribución concreta, por lo cual para poder proponer una distribución hay que proponerla completa junto con sus parámetros. Si no sabemos qué valores de los parámetros tendrá la distribución que vamos a proponer, primero debemos estimarlos.

La distribución normal tiene dos parámetros: μ y σ . Usaremos los estimadores habituales de dichos parámetros. Obtenemos:

$$\mu \cong \bar{X} = \frac{\sum_{i=1}^n x_i}{n} = 7.7256 \quad \sigma \cong S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}} = 3.1243$$

Entonces vamos a proponer que las precipitaciones son $X:N(7.7256 ; 3.1243)$. Las hipótesis nos quedan:

H_0 : Los datos recogidos provienen de una distribución normal con parámetros $\mu = 7.7256$, $\sigma = 3.1243$

H_A : Los datos recogidos no provienen de tal distribución.

Elegimos intervalos 0-0.99, 1-1.99, 2-2.99, etc. y la tabla queda:

Precipitaciones	Frecuencia	Precipitaciones	Frecuencia
$0 \leq X < 1$	1	$9 \leq X < 10$	6
$1 \leq X < 2$	2	$10 \leq X < 11$	6
$2 \leq X < 3$	1	$11 \leq X < 12$	2
$3 \leq X < 4$	1	$12 \leq X < 13$	1
$4 \leq X < 5$	1	$13 \leq X < 14$	0
$5 \leq X < 6$	3	$14 \leq X < 15$	1
$6 \leq X < 7$	7	$15 \leq X < 16$	0
$7 \leq X < 8$	6	$16 \leq X < 17$	0
$8 \leq X < 9$	2	...	0

Agrupamos algunos intervalos para que no quede ninguno con frecuencia, controlamos que ninguno quede con frecuencia extremadamente pequeña, y calculamos las probabilidades de cada intervalo (para lo cual debemos estandarizar y usar la tabla de la normal estándar). La tabla queda:

Precipitaciones	Frecuencia	Probabilidad
$X < 1$	1	0.01567
$1 \leq X < 2$	2	0.01776
$2 \leq X < 3$	1	0.03177
$3 \leq X < 4$	1	0.05134
$4 \leq X < 5$	1	0.07496
$5 \leq X < 6$	3	0.09887
$6 \leq X < 7$	7	0.11781
$7 \leq X < 8$	6	0.12682
$8 \leq X < 9$	2	0.12333
$9 \leq X < 10$	6	0.10836
$10 \leq X < 11$	6	0.08601
$11 \leq X < 12$	2	0.06167
$12 \leq X < 13$	1	0.03995
$13 \leq X$	1	0.04569

$$\chi^2 = \sum_{i=1}^k \frac{(n p_i - x_i)^2}{n p_i} = 10,979$$

Calculamos:

$\alpha = 0,05$; $k = 14$; $c = 1 + 2 = 3$, porque se estimaron 2 parámetros.

$$\chi_{0,95;11}^2 = 19,675$$

Buscamos en la tabla:

$$X^2 < \chi_{1-\alpha;k-c}^2$$

Como , no rechazamos H_0 , y por lo tanto con un nivel de significación del 5% decimos que los datos recogidos efectivamente provienen de una distribución normal.

Este material se encuentra en etapa de corrección y no deberá ser

considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 12 de julio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 23 de julio de 2004

CAPÍTULO XI

Estimación bayesiana

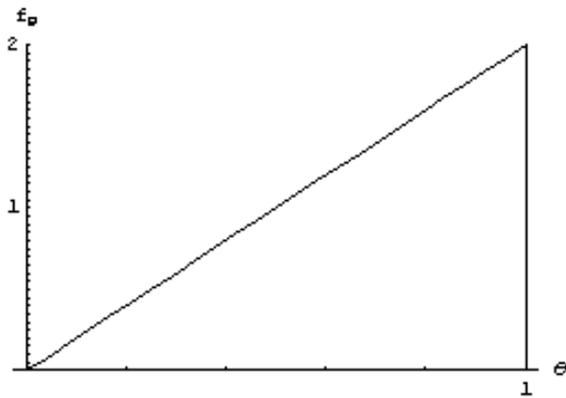
En los capítulos 8, 9 y 10 estudiamos lo que se conoce como estadística clásica. Hay, sin embargo, otra filosofía de la estadística, denominada estadística bayesiana.

La estadística bayesiana tiene la característica de que permite utilizar en la estimación de un parámetro el conocimiento previo que se tenga acerca del parámetro. En la estadística clásica, toda la información se obtiene de la muestra, y antes de tomar la muestra no se sabe nada. En la estadística bayesiana, se puede tener en cuenta para la estimación, además de los valores de la muestra, otra información subjetiva que se conozca previamente.

Por ejemplo, al estimar el parámetro p de una distribución binomial, si sabemos que el valor de p desconocido que estamos estimando está más cerca del uno que del cero, la estadística bayesiana nos permitirá incorporar dicho conocimiento a la estimación. Eso influirá en el resultado de la estimación, por lo cual la estimación bayesiana resulta sumamente controvertida. Es sumamente polémico que nuestro conocimiento subjetivo acerca del parámetro pueda influir en la estimación.

La forma de incorporar ese conocimiento previo en la estimación es comenzando por considerar que el parámetro desconocido no es una constante (como indica la estadística clásica) sino que es una **variable aleatoria**, teniendo consecuentemente una distribución de probabilidad.

Entonces, antes de tomar la muestra, proponemos una distribución para el parámetro θ . Esa distribución, que se nota f_{θ} , se denomina "**distribución a priori**", porque describe nuestro conocimiento subjetivo **antes** de tomar la muestra. Como podemos elegir arbitrariamente la distribución a priori, en esa elección tenemos la oportunidad de plasmar la información que tengamos acerca del parámetro.



Retomando el ejemplo de la estimación de un parámetro p que sabemos que es cercano a uno, podríamos elegir como distribución a priori una distribución triangular apoyada en el 1:

$$f_{\theta} = \begin{cases} 2\theta & 0 < \theta < 1 \\ 0 & \forall \text{ otro } \theta \end{cases}$$

Vemos además que dicha distribución tiene valores no nulos solamente para θ entre cero y uno, con lo cual también está incorporando nuestro conocimiento de que el parámetro p es un número entre cero y uno, y cualquier otro valor es imposible.

Luego de fijada la distribución a priori, se toma la muestra. La misma consta de los n valores $x_1 \dots x_n$.

Seguidamente, se construye la función de verosimilitud de la muestra, que como las n variables X_i son independientes, consiste en el producto de las n funciones de densidad de las X_i dado el parámetro:

$$f_{M/\theta} = f_{X_1/\theta} \dots f_{X_n/\theta} = \prod_{i=1}^n f_{X_i/\theta}$$

Lo que estamos buscando es la distribución del parámetro luego de tomar en cuenta la muestra. Es decir, la distribución "**a posteriori**" $f_{\theta/M}$. Mediante la definición de distribución condicional podemos escribir:

$$f_{\theta/M} = \frac{f_{M/\theta} f_{\theta}}{f_M}$$

Notemos que lo que está en el numerador es la distribución conjunta $f_{M\theta}$. Es la distribución conjunta de dimensión $n+1$ de las variables $\theta, X_1, X_2, \dots, X_n$. El denominador se obtiene marginando el numerador para eliminar la variable θ de la distribución conjunta. Es decir:

$$f_M = \int f_{M\theta} d\theta = \int f_{M/\theta} f_{\theta} d\theta$$

Vemos que podemos calcular el denominador a partir del numerador. Sin embargo, como el denominador no depende de θ , lo podemos ver como una simple constante que divide al numerador para que la integral de $f_{\theta/M}$ cierre a uno. Luego, escribiremos simplemente:

$$f_{\theta/M} = k f_{M/\theta} f_{\theta}$$

donde k es una constante para que la expresión resulte efectivamente ser una función de densidad (o sea que cierre a uno).

Obtuvimos de esa forma la distribución de θ dada la muestra, que luego podremos usar para estimar el valor de θ , por ejemplo tomando la esperanza de la distribución obtenida.

La estimación bayesiana permite obtener buenos resultados aunque la muestra sea chica, porque al elegir la distribución a priori podemos guiar la "tendencia" del valor del parámetro. Si la distribución que elegimos a priori es buena y refleja la realidad, entonces estamos partiendo de valores cercanos al real, y la estimación necesitará menos muestras para "aprender" la ubicación del parámetro, y por lo tanto con igual tamaño de muestra se puede llegar a resultados más precisos.

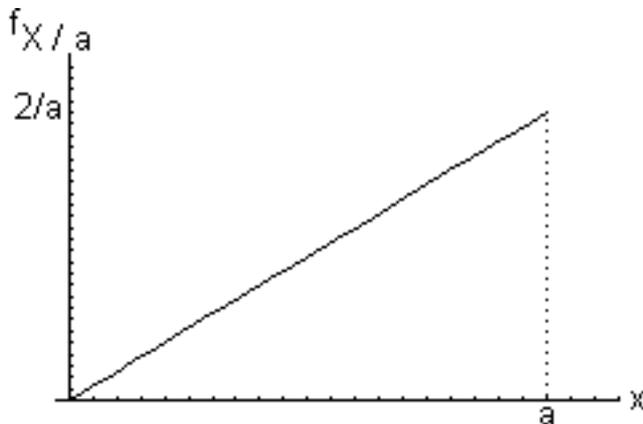
En resumen, las principales características de la estimación bayesiana son:

- El parámetro desconocido a estimar no es una constante sino una variable aleatoria.
- La estimación bayesiana permite usar el conocimiento subjetivo previo que se tenga acerca del parámetro, mediante la elección arbitraria de la distribución a priori f_{θ} .
- No nos quedan valores imposibles en la estimación, ya que podemos darle a la distribución a priori valores de densidad de probabilidad no nulos solamente para los valores posibles del parámetro.
- La estimación bayesiana permite obtener buenos resultados con muestras más pequeñas, porque la distribución no tiene que aprender la tendencia desde el comienzo, sino que puede partir con una tendencia preestablecida. Si la distribución a priori elegida es buena, entonces se convergirá al resultado más rápidamente con una menor cantidad de valores.
- La distribución del parámetro luego de tomar en cuenta la muestra se conoce como distribución a posteriori, y se obtiene de:

$$f_{\theta/M} = k f_{M/\theta} f_{\theta}$$

- Todo esto vale también para distribuciones discretas, usando donde corresponda función de probabilidad y sumatoria en vez de función de densidad e integral.
- Para algunas distribuciones en particular se sabe cuál es la distribución a priori que conviene tomar para el parámetro. Esos casos se desarrollan en las siguientes secciones de este capítulo.

Ejemplo



Se desea estimar bayesianamente el parámetro "a" de la distribución:

$$f_{X/a}(x) = \begin{cases} \frac{2x}{a^2} & 0 < x < a \\ 0 & \forall \text{ otro } x \end{cases}$$

Se sabe que "a" está entre 2 y 6. Se toma una muestra de tamaño 10, obteniéndose los valores: 2.83, 2.35, 4.88, 4.39, 3.18, 4.13, 2.23, 4.32, 2.58, 2.36

Resolución

Como sabemos que "a" está entre 2 y 6, pero no tenemos ninguna otra información, pondremos a priori para "a" una distribución uniforme entre 2 y 6:

$$f_a = \begin{cases} \frac{1}{4} & 2 < a < 6 \\ 0 & \forall \text{ otro } a \end{cases}$$

A continuación escribimos la función de densidad de la muestra dado a:

$$f_{M/a} = f_{x_1/a} \dots f_{x_n/a} = \prod_{i=1}^n f_{x_i/a} = \prod_{i=1}^n \frac{2x_i}{a^2} = \frac{2^n}{a^{2n}} \prod_{i=1}^n x_i$$

lo cual es válido para el dominio $0 < x_1 < a \cap \dots \cap 0 < x_n < a$.

Luego escribimos la distribución conjunta de la muestra con el parámetro:

$$f_{M/a} f_a = \left(\frac{2^n}{a^{2n}} \prod_{i=1}^n x_i \right) \frac{1}{4} = \frac{2^{n-2}}{a^{2n}} \prod_{i=1}^n x_i$$

lo cual es válido para el dominio $0 < x_1 < a \cap \dots \cap 0 < x_n < a \cap 2 < a < 6$.

Vemos que para que se cumplan todas las inecuaciones, a debe ser mayor que todas las x_i . Luego, simplificando, el dominio nos queda: $\max[x_i] < a < 6$.

A la expresión hallada para $f_{M/a} f_a$ la tendríamos que dividir por f_M para obtener $f_{a/M}$, pero como se dijo antes, no hace falta hacerlo ya que f_M no depende de a y entonces puede ser visto como una mera constante multiplicativa destinada a que la

integral de $f_{a/M}$ cierre a 1.

De hecho, otra observación que podemos hacer con respecto a la expresión hallada es que tanto 2^{n-2} como la productoria no dependen de a , y aparecen multiplicando a a^{-2n} . Luego, podemos quedarnos solamente con a^{-2n} considerando al resto una constante multiplicativa que no depende de a y solamente sirve para que la función de densidad cierre a uno. Es decir:

$$f_{M/a} f_a = \frac{2^{n-2}}{a^{2n}} \prod_{i=1}^n x_i = c \frac{1}{a^{2n}} = c a^{-2n}$$

En conclusión, la función de densidad a posteriori que nos queda para a es:

$$f_{a/M} = k f_{M/a} f_a = k c a^{-2n} = k' a^{-2n}$$

(juntamos las constantes k y c en una única constante k').

Ahora escribimos formalmente la función de densidad a posteriori:

$$f_{a/M} = \begin{cases} k' a^{-2n} & \max_{i=1, \dots, n} [x_i] < a < 6 \\ 0 & \forall \text{ otro } a \end{cases}$$

Con respecto a la muestra, el tamaño es 10, es decir, $n = 10$, y el máximo de la muestra vale:

$$\max_{i=1, \dots, n} [x_i] = 4.88$$

Queda:

$$f_{a/M} = \begin{cases} k' a^{-20} & 4.88 < a < 6 \\ 0 & \forall \text{ otro } a \end{cases}$$

Sólo falta integrar para encontrar el valor de k' :

$$\int f_{a/M} da = \int_{4.88}^6 k' a^{-20} da = 1 \Rightarrow k' = 2.33 \cdot 10^{14}$$

Finalmente, queda:

$$f_{a/M} = \begin{cases} 2.33 \cdot 10^{14} a^{-20} & 4.88 < a < 6 \\ 0 & \forall \text{ otro } a \end{cases}$$

Estimación puntual

Una vez obtenida la distribución a posteriori de θ , podemos estimar el verdadero valor de θ de diferentes formas, por ejemplo dando un valor. Eso se denomina estimación puntual y fue estudiado en el capítulo 8. La estimación puntual bayesiana consiste por lo general en tomar como estimación del parámetro la esperanza de la distribución a posteriori. El estimador puntual bayesiano del parámetro θ se nota θ^* .

Ejemplo

Con los datos del ejemplo anterior, la estimación puntual resulta ser:

$$\alpha^* = E(\alpha / M) = \int_{-\infty}^{\infty} \alpha \phi_{a/M} \delta \alpha = \int_{4.88}^{\infty} \alpha \cdot 2.33 \cdot 10^{14} \cdot \alpha^{-20} \delta \alpha = 5.13$$

Intervalo de confianza

El intervalo de confianza bayesiano para la estimación de un parámetro está determinado por los límites L_1 y L_2 que dejan un área de $\alpha/2$ a la izquierda y la derecha de la distribución a posteriori respectivamente, donde $1 - \alpha$ es el nivel de confianza.

Ejemplo

Con los datos del ejemplo, el intervalo del 90% de confianza ($\alpha = 0.1$) para estimar "a" viene dado por L_1 y L_2 , tales que:

$$\int_{4.88}^{L_1} 2.33 \cdot 10^{14} \cdot a^{-20} da = 0.05$$

$$\int_{L_2}^{\infty} 2.33 \cdot 10^{14} \cdot a^{-20} da = 0.05$$

Haciendo las cuentas, nos queda que $L_1 = 4.893$ y $L_2 = 5.618$

Luego, el intervalo del 90% de confianza para estimar "a" es (4.893 ; 5.618).

Problemas típicos

1) Se desea estimar bayesianamente el parámetro "a" de la distribución $X:U(a;4)$. Para eso se toma una muestra de tamaño 3, obteniéndose los valores 0.5, 1, 2. Se sabe que a es un valor positivo y cercano a cero. Haga una estimación puntual de a, y dé también un intervalo del 95% de confianza.

Resolución

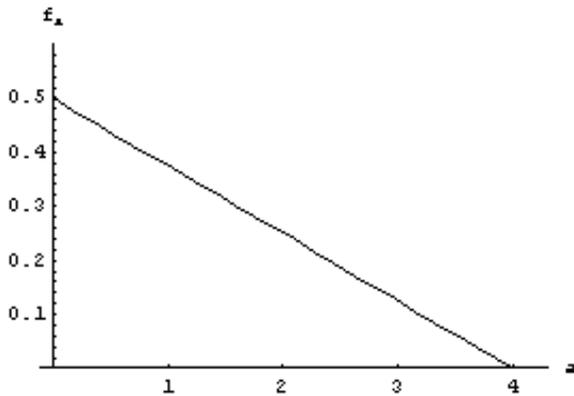
La distribución cuyo parámetro "a" queremos estudiar es:

$$f_{X/a}(x) = \begin{cases} \frac{1}{4-a} & a < x < 4 \\ 0 & \forall \text{ otro } x \end{cases}$$

Tenemos que proponer una distribución a priori para el parámetro "a". Por alguna razón sabemos que "a" es un valor positivo y cercano a cero. Observando la distribución de X, notamos también que los valores de "a" deben ser menores que

4.

Entonces nuestro conocimiento previo sobre "a" se resume así: a estará entre el cero y el cuatro, y probablemente cerca del cero.



Con dicho conocimiento previo, proponemos la siguiente distribución a priori para "a":

$$f_a = \begin{cases} \frac{4-a}{8} & 0 < a < 4 \\ 0 & \forall \text{ otro } a \end{cases}$$

A continuación, notaremos que los valores de la muestra que vamos a tomar son las variables aleatorias $X_1 \dots X_n$ distribuidas según:

$$f_{X_i}(x) = \begin{cases} \frac{1}{4-a} & a < x_i < 4 \\ 0 & \forall \text{ otro } x_i \end{cases}$$

Luego, la distribución de la muestra dado "a" es:

$$f_{M/a} = f_{X_1/a} \dots f_{X_n/a} = \frac{1}{(4-a)^n}$$

lo cual es válido para $a < x_1 < 4$, $a < x_2 < 4$, etc. Es decir, $a < \min(x_i)$.

La distribución conjunta $f_{M,a}$ es:

$$f_{M,a} = f_{M/a} f_a = \frac{1}{(4-a)^n} \frac{4-a}{8} = \frac{1}{8(4-a)^{n-1}}$$

La distribución a posteriori para "a" es:

$$f_{a/M} = k f_{M/a} f_a = \frac{k}{8(4-a)^{n-1}}$$

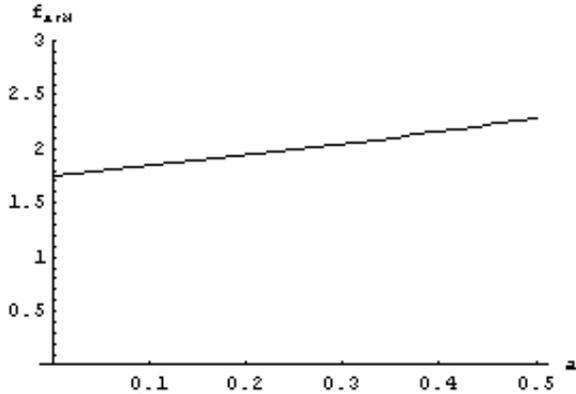
El dominio de esa función será la intersección de las condiciones de $f_{M,a}$, es decir, a

$< \min(x_i)$, y las condiciones de f_a , es decir, $0 < a < 4$. Como la distribución de las X_i va entre el cero y el cuatro, $\min(x_i) < 4$.

En resumen, el dominio resultante es $0 < a < \min(x_i)$.

Como el mínimo de la muestra es 0.5, $\min(x_i) = 0.5$.

Queda:



$$f_{a/M} = \begin{cases} \frac{k}{8(4-a)^2} & 0 < a < 0.5 \\ 0 & \forall \text{ otro } a \end{cases}$$

Haciendo la cuenta para que la integral cierre a uno, obtenemos que $k = 224$.

Luego nuestro resultado final es:

$$f_{a/M} = \begin{cases} \frac{28}{(4-a)^2} & 0 < a < 0.5 \\ 0 & \forall \text{ otro } a \end{cases}$$

Ahora vamos a hacer la estimación puntual:

$$a^* = E(a/M) = \int_{-\infty}^{\infty} a f_{a/M} da = \int_0^{0.5} \frac{28a}{(4-a)^2} da = 0.261$$

Y ahora vamos a calcular el intervalo del 90% de confianza:

$$\int_0^{L_1} \frac{28}{(4-a)^2} da = 0.05$$

$$\int_{L_2}^{0.5} \frac{28}{(4-a)^2} da = 0.05$$

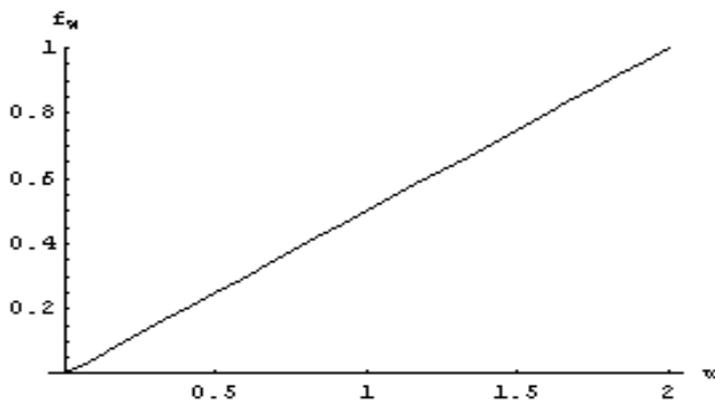
Haciendo las cuentas, nos queda que $L_1 = 0.0284$ y $L_2 = 0.4780$

Luego, el intervalo del 90% de confianza para estimar "a" es (0.0284 ; 0.4780).

2) La variable aleatoria discreta X tiene la siguiente distribución:

$$P_{X/w}(x) = \begin{cases} \frac{2-w}{3} & x = 1 \\ \frac{w}{2} & x = 2 \\ \frac{1}{3} - \frac{w}{6} & x = 3 \\ 0 & \forall \text{ otro } x \end{cases}$$

donde w es un número real entre 0 y 2. Se sabe además que es más probable que w se encuentre cerca del 2 que del 0. Para estimar bayesianamente el parámetro w , se toma una muestra de 4 valores de X obteniéndose: 2, 1, 3, 2. Haga una estimación puntual de w , y determine también un intervalo del 95% de confianza.



Resolución

Como w es un número real entre 0 y 2, y además es más probable que w se encuentre cerca del 2 que del 0, podemos plantear, por ejemplo, la siguiente distribución a priori:

$$f_w(w) = \begin{cases} \frac{w}{2} & 0 < w < 2 \\ 0 & \forall \text{ otro } a \end{cases}$$

A continuación, notaremos que los valores de la muestra que vamos a tomar son las variables aleatorias $X_1 \dots X_n$ distribuidas según:

$$P_{X_i/w}(x_i) = \begin{cases} \frac{2-w}{3} & x_i = 1 \\ \frac{w}{2} & x_i = 2 \\ \frac{1}{3} - \frac{w}{6} & x_i = 3 \\ 0 & \forall \text{ otro } x_i \end{cases}$$

La distribución de la muestra dado w es:

$$P_{M/w} = P_{X_1/w} \cdots P_{X_n/w}$$

Usando la muestra, podemos evaluar las $P_{X_i/w}$ en los correspondientes x_i :

$$\begin{aligned} P_{M/w} &= P_{X_1/w}(x_1)P_{X_2/w}(x_2)P_{X_3/w}(x_3)P_{X_4/w}(x_4) = \\ &= P_{X_1/w}(2)P_{X_2/w}(1)P_{X_3/w}(3)P_{X_4/w}(2) = \frac{w}{2} \frac{2-w}{3} \left(\frac{1}{3} - \frac{w}{6} \right) \frac{w}{2} \end{aligned}$$

Simplificando un poco se obtiene:

$$P_{M/w} = \frac{w^2}{18} \left(1 - w + \frac{w^2}{4} \right)$$

La distribución conjunta de la muestra con el parámetro es:

$$P_{M/w} f_w = P_{M/w} \frac{w}{2} = \frac{w^3}{36} \left(1 - w + \frac{w^2}{4} \right)$$

La distribución a posteriori es:

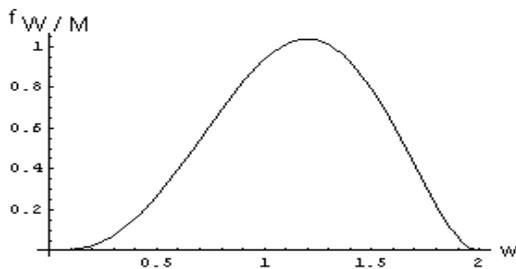
$$f_{w/M} = k P_{M/w} f_w$$

Es decir:

$$f_{w/M} = \begin{cases} \frac{k}{36} w^3 \left(1 - w + \frac{w^2}{4} \right) & 0 < w < 2 \\ 0 & \forall \text{ otro } w \end{cases}$$

Hallamos el valor de k:

$$\int_0^2 f_{w/M} da = \int_0^2 \frac{k}{36} w^3 \left(1 - w + \frac{w^2}{4} \right) dw = 1 \Rightarrow k = 135$$



Luego, la distribución a posteriori es:

$$f_{w/M} = \begin{cases} \frac{15}{4} w^3 \left(1 - w + \frac{w^2}{4} \right) & 0 < w < 2 \\ 0 & \forall \text{ otro } w \end{cases}$$

Hacemos la estimación puntual:

$$w^* = E(W / M) = \int_{-\infty}^{\infty} w f_{W/M} dw = \int_0^1 w \frac{15}{4} w^3 \left(1 - w + \frac{w^2}{4}\right) da = 1.14$$

Hallamos el intervalo del 95% de confianza:

$$\int_0^{L_1} \frac{15}{4} w^3 \left(1 - w + \frac{w^2}{4}\right) da = 0.025$$

$$\int_{L_2}^1 \frac{15}{4} w^3 \left(1 - w + \frac{w^2}{4}\right) da = 0.025$$

Haciendo las cuentas, nos queda que $L_1 = 0.446$ y $L_2 = 1.764$

Luego, el intervalo del 95% de confianza para estimar w es $(0.446 ; 1.764)$.

3) Una máquina produce piezas cuyo peso está dado por la siguiente distribución:

$$f_{x/\beta}(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

La máquina tiene una perilla que permite seleccionar el valor de β . Pero un día la perilla se rompe, y el operario no sabe si quedó ajustada en $\beta = 5$ ó en $\beta = 8$. Sabe que una de las dos es correcta, pero no está seguro de cuál.

Mirando la perilla rota, le parece que $\beta = 8$ es el doble de probable que $\beta = 5$. Para sacarse la duda, toma una muestra de 6 piezas producidas por la máquina y obtiene los siguientes pesos: 17.22, 3.49, 9.57, 1.36, 0.91, 20.86. ¿Qué puede informar?

Resolución

Si las únicas dos opciones son $\beta = 5$ y $\beta = 8$, y además $\beta = 8$ es el doble de probable que $\beta = 5$, entonces se puede fijar la siguiente distribución a priori para el parámetro β :

$$P_{\beta}(\beta) = \begin{cases} 1/3 & \beta = 5 \\ 2/3 & \beta = 8 \\ 0 & \forall \text{ otro } \beta \end{cases}$$

Los pesos de las piezas que toma el operario son las variables aleatorias $X_1 \dots X_n$ distribuidas según:

$$f_{x_i/\beta}(x_i) = \begin{cases} \frac{1}{\beta} e^{-\frac{x_i}{\beta}} & x_i > 0 \\ 0 & x_i \leq 0 \end{cases}$$

La distribución de la muestra dado β es:

$$f_{M/\beta} = f_{x_1/\beta} \cdots f_{x_n/\beta} = \frac{1}{\beta^n} \prod_{i=1}^n e^{-\frac{x_i}{\beta}}$$

La distribución conjunta de la muestra con el parámetro es:

$$f_{M/\beta} f_{\beta} = \begin{cases} \frac{1}{3\beta^n} \prod_{i=1}^n e^{-\frac{x_i}{\beta}} & \beta = 5 \\ \frac{2}{3\beta^n} \prod_{i=1}^n e^{-\frac{x_i}{\beta}} & \beta = 8 \\ 0 & \forall \text{ otro } \beta \end{cases}$$

La distribución a posteriori para β es:

$$f_{\beta/M} = k f_{M/\beta} f_{\beta} = \begin{cases} k \frac{1}{3\beta^n} \prod_{i=1}^n e^{-\frac{x_i}{\beta}} & \beta = 5 \\ k \frac{2}{3\beta^n} \prod_{i=1}^n e^{-\frac{x_i}{\beta}} & \beta = 8 \\ 0 & \forall \text{ otro } \beta \end{cases}$$

Ahora vamos a usar la información de la muestra. Ponemos $n = 6$ y reemplazamos los x_i por los valores obtenidos.

i	x_i	$e^{-x_i/5}$	$e^{-x_i/8}$
1	17.22	0.0319	0.1162
2	3.49	0.4976	0.6465
3	9.57	0.1475	0.3023
4	1.36	0.7619	0.8437
5	0.91	0.8336	0.8925
6	20.86	0.0154	0.0737

$$\frac{1}{3} \frac{1}{5^n} \prod_{i=1}^n e^{-\frac{x_i}{5}} = \frac{1}{3} \frac{1}{5^6} 2.29544 \cdot 10^{-5} = 4.89694 \cdot 10^{-10}$$

$$\frac{2}{3} \frac{1}{8^n} \prod_{i=1}^n e^{-\frac{x_i}{8}} = \frac{2}{3} \frac{1}{8^6} 0.0012605 = 3.2056 \cdot 10^{-9}$$

Para que la distribución cierre a 1, $k = 2.7061 \cdot 10^8$. Queda:

$$f_{\beta/M} = \begin{cases} 0.1325 & \beta = 5 \\ 0.8675 & \beta = 8 \\ 0 & \forall \text{ otro } \beta \end{cases}$$

La opción $\beta = 8$ quedó mucho más probable que la opción $\beta = 5$. Entonces la opción correcta es probablemente $\beta = 8$.

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 23 de julio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 25 de julio de 2004_

Algunos parámetros particulares

En la sección anterior se estudió la manera de estimar bayesianamente un parámetro cualquiera de una distribución cualquiera. También se estableció que como distribución a priori se puede usar por lo general cualquier distribución que se desee.

Sin embargo, para algunos parámetros en particular, de algunas distribuciones en particular, **conviene** tomar determinadas distribuciones como distribuciones a priori. Esto no constituye una contradicción con el párrafo anterior, porque no estamos diciendo que no se pueda tomar cualquier distribución como distribución a priori, sino que simplemente en algunos casos se sabe que **conviene** tomar una determinada distribución.

Para ilustrar las ventajas de tomar una distribución a priori conveniente, se analizarán en detalle la estimación del parámetro p de una distribución de Bernoulli y del parámetro p de una distribución Binomial. Luego se dará una tabla que contempla el resto de los casos.

El parámetro p de una distribución de Bernoulli

Para estimar el parámetro p de una distribución de Bernoulli se toma una muestra de n instancias independientes del experimento de Bernoulli, y se llama r a la cantidad de éxitos obtenidos.

Si se toma como distribución a priori una distribución Beta(a ; b), entonces la distribución a posteriori será Beta($a+r$; $b+n-r$).

Comentarios:

- Vemos que no tendremos que hacer prácticamente ninguna cuenta para obtener la distribución a posteriori.
- La ventaja de que la posteriori nos vuelva a dar beta, es que la podemos usar a su vez como la priori de la siguiente estimación, y volver a obtener beta, y así sucesivamente. Esto permite serializar las estimaciones.
- Eso quiere decir que si luego de hecha la estimación obtenemos más información (o sea, mayor cantidad de observaciones) entonces podemos tomar como priori la posteriori que obtuvimos, y obtener una nueva posteriori incorporando la nueva información. Es decir, a medida que obtenemos más información, podemos ir *mejorando* la estimación, en vez de tener que volver a hacerla desde el principio.

- Por último, observemos que da lo mismo hacer la estimación con toda la información al mismo tiempo, que ir haciendo estimaciones incorporando la información por partes. El resultado final será el mismo. La utilidad de este método es que podemos ir usando la información a medida que la vamos obteniendo.

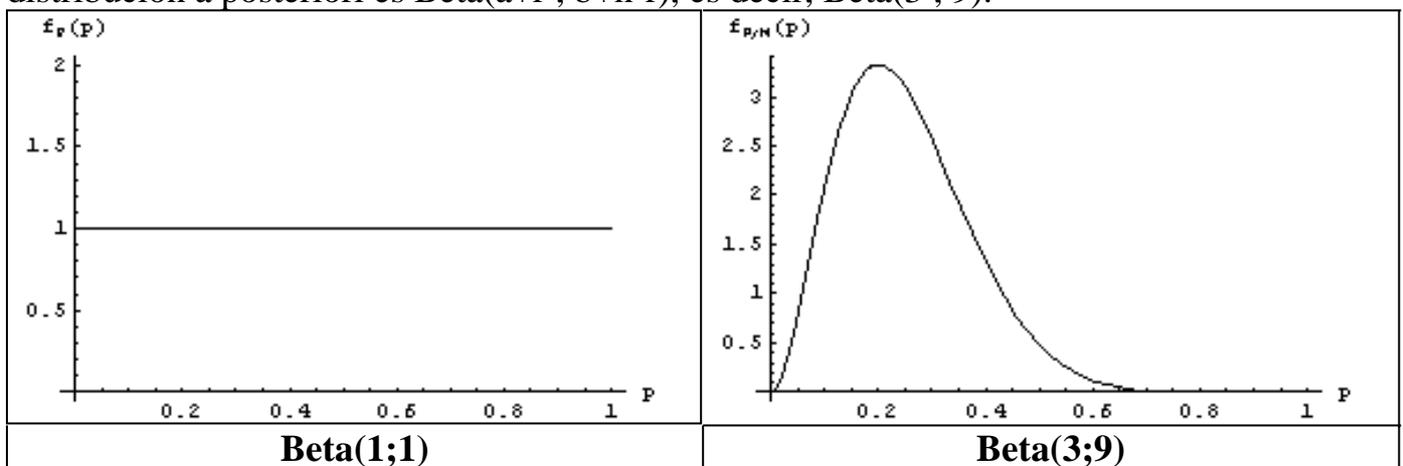
Con respecto a la distribución Beta en sí, fue estudiada en la sección 7 del capítulo 7 (se recomienda repasar dicha sección). Sus características la hacen ideal como distribución para p , porque sus valores van entre 0 y 1 y sus parámetros le permiten adquirir las más diversas formas, teniendo una gran flexibilidad.

Ejemplo 1

Se desea estimar el parámetro p de una distribución de Bernoulli. No se tiene ninguna información sobre p . Se toma una muestra de tamaño 10, obteniéndose 2 éxitos.

Como no se sabe nada sobre p , tendríamos que poner a priori una distribución Uniforme(0;1). Pero para estimar p es conveniente usar una distribución Beta. Sucede que cuando los parámetros a y b valen 1, la distribución Beta coincide con la Uniforme(0;1). Es decir, $Beta(1;1) \equiv Uniforme(0;1)$. Entonces planteamos como priori $Beta(1;1)$.

La muestra que tomamos nos proporciona la información $n = 10$, $r = 2$. Luego, la distribución a posteriori es $Beta(a+r ; b+n-r)$, es decir, $Beta(3 ; 9)$.



Vemos que a posteriori la distribución tendió a concentrarse hacia la izquierda, más precisamente en torno al 0.2, que constituye el punto más probable. La moda de

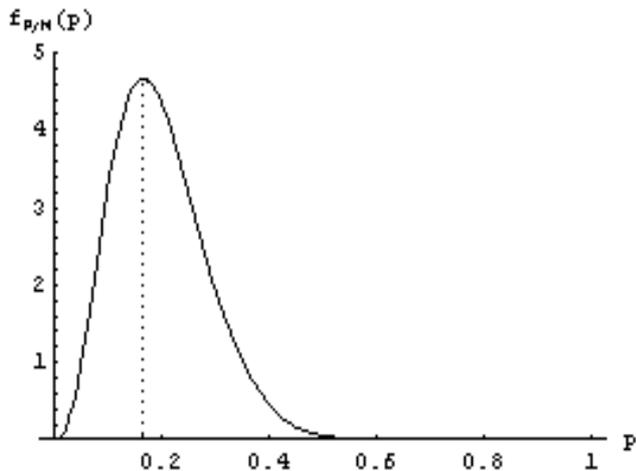
esta distribución vale $\frac{a-1}{a+b-2} = \frac{3-1}{3+9-2} = 0.2$. En la estadística clásica, nuestra

estimación habría sido $\hat{p} = \frac{r}{n} = \frac{2}{10} = 0.2$. Observamos entonces que la moda de la beta que obtenemos coincide con la estimación clásica, pero tenemos las ventajas de contar con una distribución en vez de un simple valor.

Ejemplo 2

Luego de hacer la estimación anterior, se hacen 8 observaciones más, obteniendo 1 éxito.

Ahora vamos a utilizar como priori la información que ya tenemos sobre p , es decir, $Beta(3;9)$. La nueva muestra que tomamos nos proporciona la información $n = 8$, $r = 1$. Luego, la distribución a posteriori es $Beta(3+r ; 9+n-r)$, es decir, $Beta(4 ; 16)$. Ahora el valor más probable es:



$$\frac{a-1}{a+b-2} = \frac{4-1}{4+16-2} = 0.17$$

Vemos que la distribución "asimiló" la nueva información adicionada y se desplazó un poco más a la izquierda.

Hagamos ahora algunas observaciones:

1) ¿Qué habría pasado si hubiéramos usado toda junta la información de las dos muestras, es decir, si en vez de considerar una primera muestra de tamaño 10 con 2 éxitos y una segunda muestra de tamaño 8 con 1 éxito hubiéramos considerado una única muestra de tamaño 18 con 3 éxitos?

La distribución a priori original era $Beta(1;1)$. La distribución a posteriori final sería $Beta(1+3;1+18-3)$, es decir, $Beta(4;16)$. Vemos que efectivamente habríamos obtenido el mismo resultado, pero la ventaja de proceder como lo hicimos radica en que fuimos utilizando la información a medida que la fuimos obteniendo, con lo cual no tuvimos que esperar a tener toda la información, y fuimos aprovechando la información parcial para hacer estimaciones provisionarias. Desde luego que tener estimaciones provisionarias es mejor que no tener nada.

2) ¿Después de haber encadenado dos estimaciones bayesianas sucesivas, el resultado sigue coincidiendo con el resultado clásico?

Sí, porque como dijimos en 1), usar toda la información junta da el mismo resultado que hacer las estimaciones intermedias. Luego, como el resultado clásico coincide con el resultado de usar toda la información junta, debe coincidir también

$$\hat{p} = \frac{r}{n} = \frac{3}{18} = 0.17$$

con lo que obtuvimos. En efecto,

3) ¿Qué sucede con la varianza?

La varianza de la Beta(1,1) es $\frac{1 \cdot 1}{(1+1)^2 (1+1+1)} = 0.0833$

La varianza de la Beta(3,9) es $\frac{3 \cdot 9}{(3+9)^2 (3+9+1)} = 0.0144$

La varianza de la Beta(4,16) es $\frac{4 \cdot 16}{(4+16)^2 (4+16+1)} = 0.0076$

Vemos que a medida que vamos agregando más información, la varianza va disminuyendo, es decir que la distribución se va concentrando más. Eso le va dando mayor confiabilidad a la estimación.

El parámetro p de una distribución Binomial

Para estimar el parámetro p de una distribución de Binomial(n;p) se toma muestra una muestra de tamaño m. Se llama r_i a la cantidad de éxitos del elemento i de la

$$R = \sum_{i=1}^m r_i$$

muestra, y se calcula

Si se toma como distribución a priori una distribución Beta(a ; b), entonces la distribución a posteriori será Beta(a+R ; b+mn-R).

Comentarios:

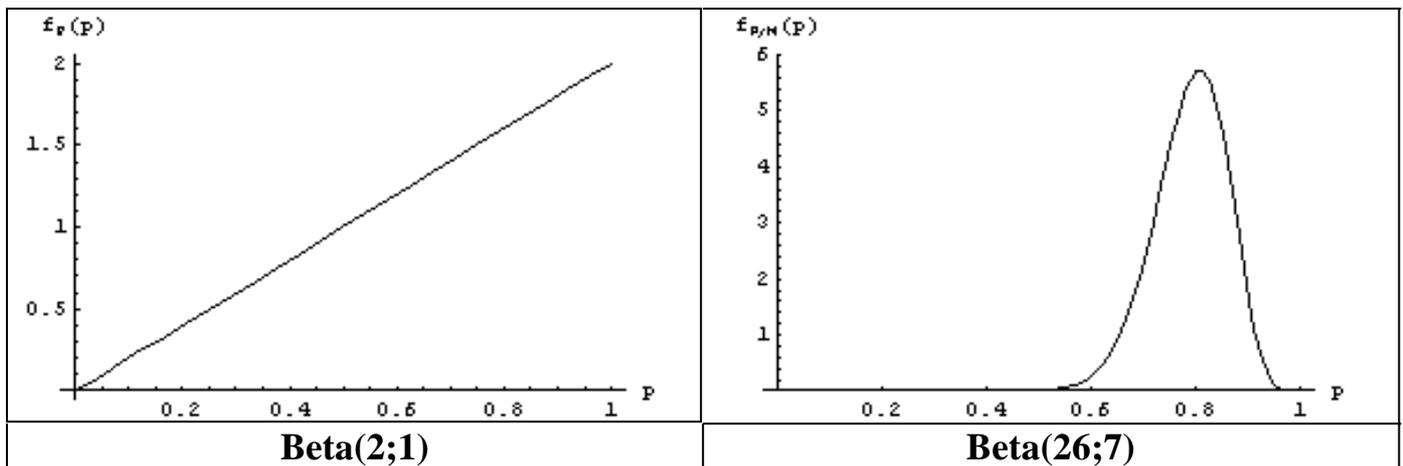
- Cada elemento de la muestra está constituido por n iteraciones del experimento de Bernoulli. Luego, se tienen en total m.n iteraciones del experimento de Bernoulli, con R éxitos en total. Vista de esa forma, esta estimación es idéntica a la del parámetro p de una distribución de Bernoulli. Por eso al parámetro "a" se le suma la cantidad de éxitos, y al parámetro "b" se le suma la cantidad de iteraciones y se le resta la cantidad de éxitos, igual que en la estimación del parámetro p de la distribución de Bernoulli.

Ejemplo

Se desea estimar el parámetro p de una distribución Binomial(n = 10, p). Se sabe además que el parámetro p está probablemente más cerca del uno que del cero. Se toman 3 muestras, obteniéndose 7, 9 y 8 éxitos.

Vamos a tomar como priori una distribución beta que asigne más probabilidad a los valores cercanos al uno que a los cercanos al cero, por ejemplo la distribución Beta(2;1) (ver capítulo 7, sección 7).

La muestra nos indica que m = 3 y R = 7 + 9 + 8 = 24. Luego, la distribución a posteriori es Beta(2+R ; 1+m.n-R), es decir Beta(26 ; 7).



La moda de la distribución Beta(22;7) es $\frac{a-1}{a+b-2} = \frac{26-1}{26+7-2} = 0.806$

La estadística clásica habría requerido usar máxima verosimilitud. Como se vio en

el capítulo 8, la estimación clásica habría sido $\hat{p} = \frac{R}{m.n} = \frac{24}{30} = 0.8$. Esto parece no coincidir, pero es solamente porque no tomamos una uniforme como priori. Si hubiéramos tomado una uniforme como priori, es decir, una Beta(1;1), la posteriori

habría sido Beta(25;7), y la moda habría sido $\frac{25-1}{25+7-2} = 0.8$. Observemos entonces que en la estimación clásica estamos sin darnos cuenta tomando una uniforme como priori. Y eso, en estimación bayesiana, es simplemente un caso particular.

Distribuciones convenientes

Distribución y parámetro a estimar	Muestra y cálculos	Distribución a priori	Distribución a posteriori
Bernoulli(p) estimar p	n: tamaño de la muestra r: cantidad de éxitos	Beta(a;b)	Beta(a+r ; b+n-r)
Binomial(n;p) estimar p	m: tamaño de la muestra $R = \sum_{i=1}^m r_i$ (cantidad total de éxitos)	Beta(a;b)	Beta(a+R ; b+m.n-R)
Geométrica(p) estimar p	m: tamaño de la muestra $N = \sum_{i=1}^m n_i$ (cantidad total de iteraciones)	Beta(a;b)	Beta(a+m ; b+N-m)

Pascal(k;p) estimar p	n: tamaño de la muestra $N = \sum_{i=1}^n n_i$ (cantidad total de iteraciones)	Beta(a;b)	Beta(a+m.k ; b+N-m.k)
Poisson(μ) estimar μ	n: tamaño de la muestra $R = \sum_{i=1}^n r_i$ (cantidad total de eventos)	Gamma(k;h)	Gamma(k+R ; h+n)
ExpNeg(λ) estimar λ	n: tamaño de la muestra $X = \sum_{i=1}^n x_i$ (tiempo total)	Gamma(k;h)	Gamma(k+n ; h+X)
Gamma(r; λ) estimar λ	n: tamaño de la muestra $X = \sum_{i=1}^n x_i$ (tiempo total)	Gamma(k;h)	Gamma(k+n.r ; h+X)
Normal(μ ; σ) estimar μ	n: tamaño de la muestra $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$ (promedio muestral)	Normal(m,d)	Normal(m',d') $m' = \frac{\sigma^2 \mu + n d^2 \bar{X}}{\sigma^2 + n d^2}$ $d' = \frac{\sigma^2 d^2}{\sigma^2 + n d^2}$

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 25 de julio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 10 de junio de 2004

APÉNDICE A

Cálculo combinatorio

El cálculo combinatorio es una herramienta matemática que, dada una determinada cantidad de elementos, permite calcular de cuántas formas posibles podemos tomar una parte de ellos y/u ordenarlos.

Por ejemplo, si tenemos un mazo de 52 cartas, y un jugador recibe 5 cartas de ese mazo, nos puede interesar calcular cuántas manos distintas podría recibir. Es decir, cuántas "combinaciones" se pueden formar con 5 cartas tomadas de entre 52.

Antes de poder hacer el cálculo, es necesario determinar algunas cosas:

- Las cantidades: debemos determinar cuántos elementos hay en total, y cuántos vamos a tomar. En el ejemplo anterior, tomamos 5 elementos de 52.
- La naturaleza: debemos determinar si estamos tomando todos los elementos disponibles, o sólo algunos de ellos. Por ejemplo, tomando 5 cartas entre 52, importará cuáles tomamos (es decir, importa la naturaleza de la selección). En cambio, si solamente nos interesa de cuántas formas podemos ordenar 5 libros, no nos interesa la naturaleza, porque no tenemos que elegir determinados libros sino que vamos a estar trabajando con los 5 al mismo tiempo.
- El orden: debemos determinar si nos interesa o no nos interesa el orden en que tomamos los elementos. Por ejemplo, si nos importa el orden, tirar un dado y sacar un 5 y luego un 3, no es lo mismo que sacar un 3 y luego un 5. Serían dos resultados distintos. En cambio si no nos importa el orden, sacar un 5 y luego un 3 ó un 3 y luego un 5 es lo mismo, y los dos casos constituirán un único resultado.
- La repetición: tiene que ver con si se puede elegir más de una vez o no el mismo elemento. Por ejemplo, si en una caja hay una bolita blanca, una negra, y una violeta, y vamos a sacar dos, si lo hacemos con reposición entonces habrá repetición, porque es posible sacar dos veces la misma bolita.

Ejemplo 1

Me gané un viaje al caribe para mí y 2 amigos. Pero tengo 5 amigos, así que voy a tener que elegir a 2. Si voy a calcular cuántas decisiones distintas podría tomar, ¿cuáles son los factores involucrados?

- Las cantidades: vamos a elegir 2 elementos de un total de 5.

- La naturaleza: los 5 elementos son todos distinguibles entre sí. Invitar a Juan no es lo mismo que invitar a Pedro. O sea: como no puedo elegir a todos, importa a cuáles elijo.
- El orden: en este caso el orden en que escoja los 2 elementos no importa. Invitar a Martín y a Nicolás es lo mismo que invitar a Nicolás y a Martín.
- Repetición: no se puede elegir dos veces al mismo amigo. Deben ser dos personas distintas.

Ejemplo 2

Una habitación tiene 4 paredes, y tengo 4 colores distintos para pintarlas. No voy a mezclar colores, y voy a pintar cada pared de un color distinto. Si voy a calcular de cuántas formas distintas puedo pintar la habitación, ¿cuáles son los factores involucrados?

- Las cantidades: vamos a usar 4 colores de un total de 4. Es decir, vamos a elegir 4 elementos de un total de 4 elementos. Vamos a usar todos los elementos.
- La naturaleza: los 5 elementos son todos distinguibles entre sí. Vamos a usar todos los elementos, así que esta decisión no es importante "cuáles elementos" elijo.
- Orden: el orden sí es importante. Observemos que si no importa cuáles elementos elegimos, lo único que va a importar es el orden en que los elijamos. Elegir el rojo para la primera pared y el verde para la segunda no es lo mismo que elegir el verde para la primera pared y el rojo para la segunda.
- Repetición: no se puede elegir dos veces el mismo color.

Ahora veremos cuáles son los modelos a los que corresponden las formas de tomar los elementos. Los modelos se pueden clasificar:

- Según si hay o no hay elementos repetidos:
 - En los modelos simples: todos los elementos son distintos (distinguibles) y se eligen todos una sola vez.
Ejemplo: a b c d e
 - En los modelos compuestos, puede haber elementos iguales (no distinguibles) o bien se puede elegir un mismo elemento más de una vez.
Ejemplo: a b b c d
- Según qué importa:
 - En las variaciones, importan la naturaleza y el orden. Es decir, importa CUÁLES elementos elijo, y EN QUÉ ORDEN.
 - En las permutaciones, importa solamente el orden. Es decir, no importa cuáles elementos elijo sino EN QUÉ ORDEN.
 - En las combinaciones, importa solamente la naturaleza. Es decir, importa CUÁLES elementos elijo pero no importa en qué orden los elijo.

Tabla rápida de consulta de fórmulas

Primero daremos sin ninguna explicación ni demostración las 6 fórmulas. Se representa mediante 'n' la cantidad total de elementos, y mediante 'k' la cantidad de elementos que se toman:

Modelos simples (sin repetición):

Modelo	Fórmula	Importa	Ejemplo
Permutación	$P_n = n!$	orden	Formas de ordenar {a,b,c}: abc, acb, bac, bca, cab, cba $P_3 = 3! = 6$
Variación	$V_{n,k} = \frac{n!}{(n-k)!}$	naturaleza ("¿cuáles?") y orden	Formas de tomar 2 elementos de {a,b,c}, teniendo en cuenta el orden: ab, ba, ac, ca, bc, cb $V_{3,2} = 3! / 1! = 6$
Combinación	$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$	naturaleza	Formas de tomar 2 elementos de {a,b,c}, sin tener en cuenta el orden: ab, ac, bc $C_{3,2} = 3! / 2!1! = 6/2 = 3$

Modelos compuestos (con repetición):

Modelo	Fórmula	Importa	Ejemplo
Permutación	$P'_{n_1, n_2, \dots, n_k} = \frac{(n_1 + n_2 + \dots + n_k)!}{n_1! n_2! \dots n_k!}$	orden	Formas de ordenar {a,a,b,c} aabc, aacb, abac, acab, abca, abca, baca, caba, baac, caab, bcaa, cbaa $P'_{2,1,1} = 4! / 2!1!1! = 24/2 = 12$
Variación	$V'_{n,k} = n^k$	naturaleza ("¿cuáles?") y orden	Formas de tomar 3 elementos de {a,b} (pudiendo repetir) y teniendo en cuenta el orden aaa, aab, aba, abb, baa, bab, bba, bbb $V'_{2,3} = 2^3 = 8$
Combinación	$C'_{n,k} = \frac{(n+k-1)!}{(n-1)!k!}$	naturaleza	Formas de tomar 3 elementos de {a,b} (pudiendo repetir) aaa, aab, abb, bbb $C'_{2,3} = 4! / 1!3! = 24/6 = 4$

A continuación nos detendremos caso por caso:

Permutación simple

Se tienen n elementos, y se desea ver de cuántas formas se los puede ordenar. Es decir, los elementos son siempre los mismos, y cada forma posible sólo difiere de

las demás en el orden en que se toman los elementos.

- Fórmula

$$P_n = n!$$

donde n es la cantidad de elementos a ordenar

- Ejemplo 1

¿De cuántas formas se pueden ordenar los elementos $\{a,b,c\}$?

abc, acb, bac, bca, cab, cba (6 formas)

$$P_3 = 3! = 6$$

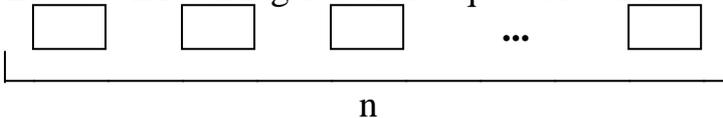
- Ejemplo 2

Se tienen 5 libros que se desea poner en un estante. ¿De cuántas formas posibles se los puede ordenar?

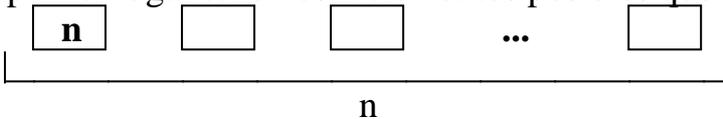
La cantidad total de formas posibles de ordenar n elementos es $P_n = n!$. Entonces la cantidad de formas posibles de ordenar los 5 libros es $5! = 120$.

- Deducción de la fórmula

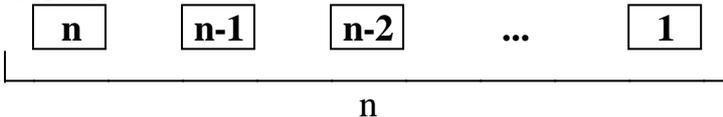
Estos son los n lugares en los que colocaremos los n elementos:



Vamos a ir colocando los elementos en los lugares de izquierda a derecha. En el primer lugar tenemos n elementos posibles que podemos colocar.



Para el segundo lugar ya nos quedarán sólo $n-1$ elementos. Para el tercero $n-2$, y así hasta que en el último (n -ésimo) lugar, sólo nos quedará un elemento posible para ubicar.



Entonces la cuenta fue $n(n-1)(n-2)(n-3)...1 = n!$

Por ejemplo, si tenemos 5 libros, para la primera posición tenemos 5 opciones, para la segunda 4, para la tercera 3, para la cuarta 2 y para la quinta 1.

$$5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5!$$

Variación simple

Es como la permutación, pero no se usan los n elementos sino que se usan solamente k de ellos. Entonces habrá que tener en cuenta no solamente el orden, sino cuáles de los n elementos se eligen (naturaleza).

- Fórmula

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

donde n es la cantidad total de elementos, y k es la cantidad de elementos que se eligen. Se lee: "variaciones de n elementos tomados de k ".

• Ejemplo 1

Se tienen los elementos {a,b,c,d}. ¿De cuántas formas se puede tomar 2 de ellos, sin repetir ninguno, y teniendo en cuenta el orden?

Comencemos por aclarar que:

- 1) tener en cuenta el orden significa que "ab" ≠ "ba"
- 2) tener en cuenta la naturaleza significa que elegir al a y al b no es lo mismo que elegir al a y al c.

Entonces las variaciones en este caso son:

ab, ba, ac, ca, ad, da, bc, cb, bd, db, cd, dc

$$V_{4,2} = 4! / 2! = 24 / 2 = 12$$

• Observación

Cuando $n = k$ (es decir, cuando se toman todos los elementos) deja de importar "cuáles" elementos se eligen, porque se están eligiendo todos, y solamente importa el orden. Y el modelo en el que sólo importa el orden es la permutación. Vemos entonces que la permutación simple es un caso particular de la variación simple. De hecho cuando $n=k$, la fórmula de la variación simple $n!/(n-k)!$ se reduce a $n!/0! = n!$, que es justamente la fórmula de la permutación simple.

• Ejemplo 2

Hay 5 participantes en un determinado concurso. El jurado debe otorgar primer premio y segundo premio. ¿Cuántas decisiones distintas puede tomar el jurado?

Es un caso de variaciones porque:

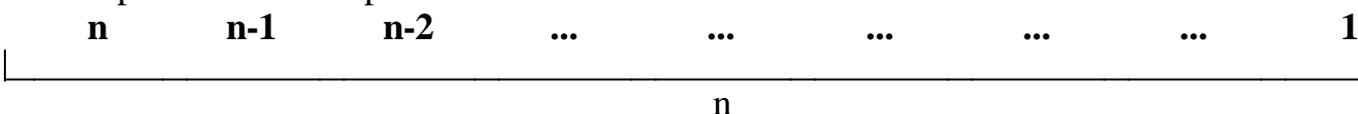
- 1) Entre 5 participantes, serán elegidos 2 (naturaleza)
- 2) Darle el primer premio a Juan y el segundo a Pedro no es lo mismo que darle el primer premio a Pedro y el segundo a Juan. (orden)

En este caso las variaciones son simples porque no se puede elegir dos veces al mismo elemento (no se le puede dar a la misma persona los dos premios)

Entonces la respuesta es $V_{5,2} = 5! / 3! = 120 / 6 = 20$

• Deducción de la fórmula

Para la permutación simple teníamos:



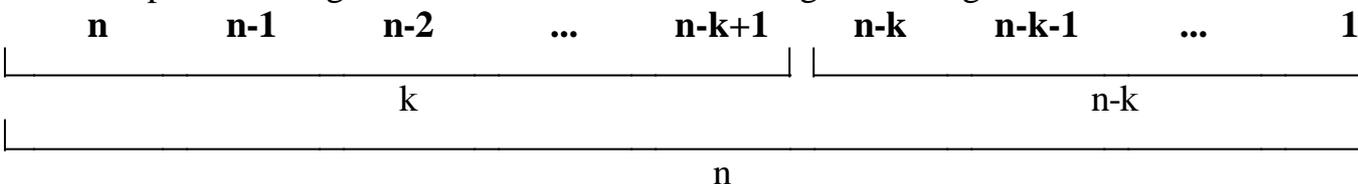
Es decir, teníamos n posiciones; para la primera posición teníamos n opciones, para la segunda n-1, etc.

Ahora tendremos solamente k posiciones. Para la primera tendremos n opciones, para la segunda, n-1, para la tercera n-2, y así sucesivamente, y para la k-ésima tendremos n-k+1 opciones. Necesitamos encontrar una forma matemática de escribir el producto:

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$$

Por propiedades del factorial sabemos que esa cuenta da $n! / (n-k)!$

También podemos llegar a ese resultado mirando el siguiente diagrama:



Nos interesa solamente lo que ocurre en las k posiciones que elegimos, así que al total [n!] hay que sacarle la parte de la derecha [(n-k)!]. En el diagrama vemos que el total es:

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1) \cdot (n-k)!$$

Si queremos hacer desaparecer el (n-k)! que no nos interesa, debemos dividir n! por (n-k)!, con lo cual obtenemos $V_{n,k} = n! / (n-k)!$

Ese (n-k)! que estamos sacando porque no nos interesa es justamente P_{n-k} , es decir, la cantidad de formas de ordenar los elementos que NO elegimos (por eso no nos interesa y hay que sacarlo).

Combinación simple

Consiste en tomar k elementos entre n que hay en total, sin importar en qué orden. Es decir, importa la naturaleza ("cuáles") pero no importa el orden. Observamos que esto es como las variaciones, pero olvidándonos del orden; las variaciones distinguen "ab" de "ba", en cambio para las combinaciones "ab" = "ba", y sólo importa el hecho de que fueron "a" y "b" los elementos elegidos.

• Fórmula

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

donde n es la cantidad total de elementos, y k es la cantidad que se toman.

• Ejemplo 1

Se tienen los elementos {a,b,c,d}. ¿Cuántas formas posibles hay de elegir 2?

Comencemos por aclarar que como son combinaciones, no tenemos en cuenta el orden, con lo cual "ab" = "ba". Además recordamos que por tratarse de combinación simple, no se puede elegir 2 veces el mismo elemento. Entonces en este caso las combinaciones son:

ab, ac, ad, bc, bd, cd.

$$C_{4,2} = 4! / 2!2! = 24/4 = 6$$

Podríamos haber obtenido lo mismo tomando el resultado del ejemplo 1 de la variación simple y tachando las formas cuyos elementos ya aparecieron en otro orden:

ab, ~~ba~~, ac, ea, ad, ~~da~~, bc, e**b**, bd, ~~db~~, cd, ~~dc~~ = ab, ac, ad, bc, bd, cd

hacer esto es como decir "me deja de importar el orden", lo cual es justamente la diferencia entre variación y combinación.

• Ejemplo 2

Me gané un viaje al caribe para mí y 2 amigos. Pero tengo 5 amigos, así que voy a tener que elegir a 2. ¿Cuántas decisiones posibles puedo tomar?

Comenzamos por observar que:

1) importa la naturaleza (no es lo mismo elegir a Juan y a Pedro que a Pablo y a Carlos).

2) no importa el orden (elegir a Juan y a Pedro es lo mismo que elegir a Pedro y a Juan)

Hasta aquí sabemos que son combinaciones. Además:

3) no se puede elegir más de una vez al mismo elemento (tengo necesariamente que invitar a dos personas distintas; no puedo invitar a Juan y a Juan).

Entonces se trata de combinaciones simples. Consecuentemente, la respuesta es:

$$C_{5,2} = 5! / 2!3! = 120/12 = 10$$

• Deducción de la fórmula

Dijimos que las combinaciones eran como las variaciones, pero dejando de tener en cuenta el orden. Para las variaciones, las 6 formas abc, acb, bac, bca, cab, cba son distintas.

Para las combinaciones, esas 6 formas son una sola. Entonces si pudiéramos determinar cuántas variaciones distintas hay por cada combinación, podríamos tomar la fórmula para las variaciones y dividirla por esa cantidad, y así obtendríamos una fórmula para las combinaciones.

Veamos: si tomamos k elementos distintos (porque k es la cantidad que se toman, tanto en las variaciones como en las combinaciones) entonces tendremos una combinación. Y la cantidad de variaciones tomando k elementos, con esos k elementos que acabamos de elegir, es la cantidad de formas en que esos elementos se pueden ordenar. Y eso son las permutaciones de los k elementos. Y como según vimos antes la cantidad de permutaciones de k elementos es k!, entonces entonces por cada combinación de k elementos hay k! variaciones. Eso es lo mismo que decir que si tomamos la cantidad de variaciones y la dividimos por k!, tenemos la cantidad de combinaciones. Es decir:

$$C_{n,k} = \frac{V_{n,k}}{P_k} = \frac{V_{n,k}}{k!} = \frac{n!}{k!(n-k)!}$$

Ese resultado se denomina "número combinatorio", y se puede expresar

simplemente $\binom{n}{k}$ en vez de $\frac{n!}{k!(n-k)!}$.

Ahora repetiremos la deducción con un ejemplo concreto:

Se tienen los elementos {a,b,c,d}. ¿Cuántas formas posibles hay de tomar 3 de ellos, sin importar el orden?

abc, abd, acd, bcd

Vemos que son 4.

Por cada una de esas 4, hay 6 (es decir, 3!) variaciones, ya que por ejemplo la combinación 'abc' es el resultado de abreviar las 6 variaciones

abc, acb, bac, bca, cab, cba

Entonces si calculamos la cantidad total de variaciones, y las dividimos por 6 (es decir, 3!), deberíamos obtener la cantidad total de combinaciones. Veamos:

$$V_{4,3} = 4! / 1! = 24$$

Luego $24 / 6 = 4$, con lo cual se verifica que el resultado obtenido es correcto.

Variación con repetición

Consiste en tomar k elementos entre n que hay en total, pudiendo elegirse más de

una vez cada elemento. Es decir, por ser variación importan la naturaleza ("cuáles") y el orden, pero además, se puede elegir más de una vez cada elemento.

- Fórmula

$$V'_{n,k} = n^k$$

- Ejemplo 1

¿Cuántas formas posibles hay de tomar 2 elementos de {a,b,c}, teniendo en cuenta el orden y pudiéndose tomar más de una vez cada uno?

Veamos: aa, ab, ac, ba, bb, bc, ca, cb, cc

$$V'_{3,2} = 3^2 = 9$$

- Ejemplo 2

Quizás el ejemplo más típico de la variación con repetición es arrojar 2 dados distinguibles. ¿Cuántos resultados posibles hay?

11 12 13 14 15 16
21 22 23 24 25 26
31 32 33 34 35 36
41 42 43 44 45 46
51 52 53 54 55 56
61 62 63 64 65 66

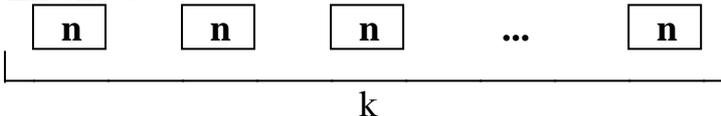
Vemos que hay 36 resultados posibles.

$$V'_{6,2} = 6^2 = 36$$

- Deducción de la fórmula

Debemos llenar k posiciones, y para cada una de ellas tenemos n opciones, porque los elementos se pueden repetir (nótese la diferencia con los modelos sin repetición, en los cuales las opciones eran n, n-1, n-2, etc).

Entonces:



Luego $V'_{n,k} = n^k$

Observemos el ejemplo 1. En la variación "aa", a pesar de ser una variación, no importa el orden, porque las 2 "a" de "aa" son iguales, pues son simplemente el mismo elemento tomado dos veces. Por eso la variación con repetición tiene la particularidad de que no en todas las formas importa el orden. Es decir, en la variación con repetición, el orden es importante solamente "cuando tiene sentido hablar de orden". Esa es la razón por la cual la fórmula de la variación con repetición es tan distinta de las otras cinco fórmulas.

Combinación con repetición

Nuevamente, la combinación es como la variación, pero sin importar el orden. Es decir, la combinación con repetición consiste en tomar k elementos de los n que hay en total (naturaleza), sin tener en cuenta el orden, y pudiendo elegir más de una vez cada elemento.

- Fórmula

$$C'_{n,k} = \frac{(n+k-1)!}{(n-1)!k!}$$

- Ejemplo 1

¿Cuántas formas posibles hay de tomar 2 elementos de {a,b,c}, sin tener en cuenta el orden y pudiéndose tomar más de una vez cada uno?

Obtenemos:

aa, ab, ac, bb, bc, cc

Podríamos haber obtenido lo mismo tomando el resultado del ejemplo 1 de la variación con repetición, y tachar las formas cuyos elementos ya aparecieron en otro orden:

aa, ab, ac, ~~ba~~, bb, bc, ea, e~~b~~, cc = aa, ab, ac, bb, bc, cc

hacer esto es como decir "me deja de importar el orden", lo cual es justamente la diferencia entre variación y combinación.

- Ejemplo 2

Hay una gran bolsa con caramelos surtidos, cuyos sabores son limón, naranja, frutilla y manzana. Nos dejan elegir dos caramelos. ¿Cuántas opciones tenemos?

Comencemos por observar que se trata de combinación porque:

1) importa la naturaleza (cuáles sabores elijo)

2) no importa el orden (elegir un caramelo de limón y uno de naranja es lo mismo que elegir un caramelo de naranja y uno de limón)

Además, es combinación con repetición porque podemos elegir, por ejemplo, dos caramelos de limón.

Entonces la respuesta es $C'_{4,2} = 5! / 3!2! = 120/24 = 5$

Permutación con repetición

Como sucedía con la permutación simple, vamos a tomar todos los elementos. Por lo tanto ya no importa la naturaleza (es decir, cuáles elementos elegimos). Importa solamente el orden. Y puede haber elementos repetidos, pero conocemos de antemano cuántos elementos hay de cada tipo. Entonces tenemos una cantidad n de elementos, que estará formada por n_1 elementos del tipo 1, n_2 elementos del tipo 2, etc. Lo que vamos a contar es todas las maneras posibles de ordenar esos elementos.

- Fórmula

$$P'_{n_1, n_2, \dots, n_k} = \frac{(n_1 + n_2 + \dots + n_k)!}{n_1! n_2! \dots n_k!}$$

- Ejemplo 1

Tenemos los elementos a, a, b, b, b. ¿De cuántas formas los podemos ordenar?

Comencemos por observar que nos vamos a ocupar solamente del orden, y que hay dos tipos de elementos, con cantidades fijas y conocidas: $n_a = 2$, $n_b = 3$.

Las permutaciones posibles son:

aaabb, ababb, abbab, abbba, baabb, babab, babba, bbaab, bbaba, bbbaa.

Vemos que hay 10. Ahora usamos la fórmula:

$$P'_{n_a, n_b} = \frac{(n_a + n_b)!}{n_a! n_b!} = \frac{5!}{2! 3!} = 10$$

• Ejemplo 2

Hay que ubicar en la puerta de la heladera 3 botellas de gaseosa, 2 de agua y una de vino. ¿De cuántas formas posibles de las puede disponer?

Comencemos por observar que se trata de permutación con repetición porque hay una cantidad fija de elementos de cada tipo y hay que calcular la cantidad de formas posibles de ordenarlos.

$$P'_{3,2,1} = \frac{(3 + 2 + 1)!}{3! 2! 1!} = \frac{6!}{3! 2! 1!} = \frac{600}{12} = 50$$

Problemas típicos

A continuación se ofrecen otros 6 problemas como complemento de los 12 ejemplos resueltos junto con las explicaciones

1) Juan tiene dos días francos por semana. ¿Cuántas formas posibles tiene el gerente de asignarle los dos francos?

Resolución

- Importa la naturaleza (importa cuáles días le asigna)
- No importa el orden (que le asigne el martes y el miércoles es lo mismo que que le asigne el miércoles y el martes)
- No hay repetición (los dos francos deben ser necesariamente días distintos)

=> Combinación simple

$$C_{7,2} = 7! / 5!2! = 21$$

2) Juan decide organizar su semana: dedicará 3 días a trabajar, 2 a estudiar y 2 a descansar. ¿Cuántas opciones tiene?

Resolución

- No importa la naturaleza (ya tiene decidido exactamente qué actividades elegir)
- Importa el orden (justamente de eso se trata este problema; no es lo mismo descansar el lunes y estudiar el martes que estudiar el lunes y descansar el martes)
- Hay repetición, y además se conocen exactamente las cantidades de veces que aparecen los elementos

=> Permutación con repetición

$$P'_{3,2,2} = 7! / 3!2!2! = 210$$

3) Juan tiene 5 calcomanías, y desea pegar una en el vidrio de adelante de su auto, y otra en el vidrio de atrás. ¿Cuántas decisiones distintas puede tomar?

Resolución

- Importa la naturaleza (importa cuáles calcomanías elige)
- Importa el orden (no es lo mismo pegar la calcomanía A en el vidrio de adelante y la calcomanía B en el vidrio de atrás, que pegar la calcomanía B en el vidrio de adelante y la calcomanía A en el vidrio de atrás)
- No hay repetición (no puede pegar dos veces la misma calcomanía; en otras palabras, tiene solamente una de cada tipo)

=> Variación simple

$$V_{5,2} = 5! / 3! = 120/6 = 20$$

4) Juan recibió 2 cartas en una determinada semana. Si le preguntan en qué día o días de esa semana recibió cartas, ¿de cuántas formas posibles puede responder?

Resolución

- Importa la naturaleza (importa en cuál o cuáles días llegaron cartas)
- No importa el orden (si la carta A le llegó el lunes y la carta B el jueves, es lo mismo que si la carta A le llegó el jueves y la carta B el lunes, puesto que lo que importa es solamente "en cuáles días recibió cartas")
- Hay repetición (las dos cartas pueden haber llegado el mismo día)

=> Combinación con repetición

$$C'_{7,2} = 8! / 6!2! = 28$$

5) Juan tiene 5 libros y desea leerlos (de a uno a la vez). ¿Cuántas opciones tiene, en cuanto al orden de lectura?

Resolución

- No importa la naturaleza (va a leer los 5 libros, así que no está eligiendo ningún grupo de ellos)
- Importa el orden (es exactamente lo que nos preguntan; no es lo mismo leer los 5 libros en el orden ABCDE que en el orden DBACE)
- No hay repetición (no leerá más de una vez el mismo libro)

=> Permutación simple

$$P_5 = 5! = 120$$

6) Juan está loco. A veces cree que es Napoleón, a veces cree que es astronauta, y a veces cree que un día lo secuestraron los marcianos mientras estaba en la ducha. Si le hacen peritajes psicológicos y le cuenta un delirio al

**doctor A y un delirio al doctor B (puede contarles a los dos el mismo delirio),
¿de cuántas formas posibles pudo delirar en los peritajes psicológicos?**

Resolución

- Importa la naturaleza ("cuáles delirios cuenta")
- Importa el orden (no es lo mismo contarle al doctor A que es Napoleón y al B que es astronauta, que contarle al doctor A que es astronauta y al B que es Napoleón)
- Hay repetición (le puede contar a los dos doctores el mismo delirio)

=> Variación con repetición

$$V'_{3,2} = 3^2 = 9$$

Los juegos de azar

A continuación se presenta un pequeño estudio de caso de algunos juegos de azar. Se hallarán determinados resultados empleando el cálculo combinatorio y la definición de probabilidad de Laplace, y luego se llegará al mismo resultado multiplicando probabilidades, para mostrar la equivalencia de los métodos.

El cálculo de la probabilidad comenzó debido a su utilidad en los juegos de azar por dinero. Es decir, el cálculo de la probabilidad se desarrolló gracias a la "timba". Es por ello que resulta frecuente encontrar en libros, guías de ejercicios, etc. ejemplos relacionados con los juegos de azar, con los que quizás el alumno no se encuentra familiarizado. Comenzaremos por explicar el significado de las expresiones más usuales:

- Honesto: un dado en el cual la probabilidad de que salga cada una de sus seis caras es $1/6$, o una moneda en la cual la probabilidad de que salga cara es $0,5$.
- Cargado: un dado o moneda no honestos.
- Naipes:
 - españoles
 - 40 cartas: 4 palos (bastos, espadas, oros y copadas). Cada palo formado por 10 cartas indicadas con el palo y un número: 1, 2, 3, 4, 5, 6, 7, 10(sota), 11(caballo) y 12(rey).
 - 50 cartas: los mismos 4 palos, con los números del 1 al 12, y además otras 2 cartas comodín.
 - ingleses
 - de póker: 52 cartas. 4 palos (diamantes, corazones, espadas y tréboles). Cada palo formado por 13 cartas con los números del 1 al 10 y las figuras J (jack), Q(reina), K(rey). El 1 es más comúnmente llamado "As".
 - 54 cartas: igual que el anterior pero incluyendo 2 payasos o comodines o jokers.
- Póker: los "juegos" se forman con 5 cartas del mazo de 52. No importa el orden en que estén las cartas.
 - Par: 2 cartas del mismo número, y las demás de otros números.
 - Par doble: 2 cartas del mismo número, otras 2 también del mismo número entre sí, y una quinta carta con un número distinto a los 2 anteriores. Ejemplo 5 5 K K 8
 - Trío o pierna: 3 cartas del mismo número, y las otras 2 de números diferentes. Ejemplo: 8 8 8 4 J.
 - Escalera: los 5 números consecutivos. El as puede ir antes del 2 o después de la K. Ejemplos: A 2 3 4 5, 4 5 6 7 8, 10 J Q K A.
 - Full house: 3 cartas del mismo número, y otras 2 del mismo número. Ejemplo: 5 5 5 J J.

- Color: las 5 cartas del mismo palo.
- Póker: 4 cartas del mismo número. Ejemplo: A A A A 7.
- Escalera real: es tener "escalera" y "color" al mismo tiempo.
- Generala: los "juegos" se forman con 5 dados. No importa el orden en que salgan los dados.
 - Escalera: 5 números consecutivos: 1-5 o 2-6.
 - Full: 3 números iguales entre sí, y otros 2 números iguales entre sí.
Ejemplo: 3 4 3 3 4
 - Póker: 4 números iguales y uno distinto. Ejemplo: 4 4 8 4 4
 - Generala: los 5 números iguales. Ejemplo: 3 3 3 3 3

A continuación veremos el cálculo de la probabilidad de cada uno de los juegos de la generala y del póker.

Los juegos de la Generala

Hay $V'_{6,5} = 7776$ resultados posibles al arrojar 5 dados. Calcularemos la probabilidad de sacar cada juego como la cantidad de formas posibles de sacar dicho juego dividido el total de resultados posibles (definición de probabilidad de Laplace).

Escalera

5 números consecutivos

Las únicas posibilidades son "1 2 3 4 5" y "2 3 4 5 6" (obviamente, en cualquier orden).

Por cálculo combinatorio:

Escaleras posibles	Formas de ordenar
Vamos a tomar 1 entre 2 escaleras posibles	son las formas de ordenar 5 dados distinguibles entre sí
$C_{2,1}$	P_5

Queda $C_{2,1} \cdot P_5 = 240$ resultados en un total de $V'_{6,5}$ resultados posibles.

$\Rightarrow P(\text{escalera}) = 0,0308641975$

Multiplicando probabilidades:

Vemos que para obtener escalera hay que sacar obligatoriamente un 2, un 3, un 4, un 5, y además un 1 o un 6.

Voy a sacar 2, 3, 4, 5, [1 ó 6] y luego lo voy a desordenar.

- Primero tengo 1/6 de números favorables (sacar un 2)
- Luego tengo 1/6 de números favorables (sacar un 3)
- Luego tengo 1/6 de números favorables (sacar un 4)

- Luego tengo 1/6 de números favorables (sacar un 5)
 - Luego tengo 2/6 números favorables (sacar un 1 ó un 6)
 - Hay 120 maneras posibles de ordenarlo ($P_5 = 5! = 120$).
- Queda $1/6 \cdot 1/6 \cdot 1/6 \cdot 1/6 \cdot 2/6 \cdot 120 = 0,0308641975$
 $\Rightarrow P(\text{full}) = 0,0308641975$

Full

3 números iguales entre sí, y otros 2 números iguales entre sí.
 Ejemplo: 3 4 3 3 4

Por cálculo combinatorio:

Números posibles	Formas de ordenar
vamos a usar 2 números de un total de 6, y es importante cuál número será para el trío y cuál para el par (es decir, importa el orden).	tenemos para ordenar 3 elementos indistinguibles entre sí y otros 2 elementos indistinguibles entre sí.
$V_{6,2}$	$P'_{3,2}$

Queda $V_{6,2} \cdot P'_{3,2} = 449280$ resultados en un total de $V'_{6,5}$ resultados posibles.
 $\Rightarrow P(\text{full}) = 0,0385802469$

Multiplicando probabilidades:

Voy a sacar $a a a b b$ y luego lo voy a desordenar.

- Primero tengo 6/6 de números favorables (saco un número cualquiera)
- Luego tengo 1/6 de números favorables (saco el mismo número)
- Luego tengo 1/6 de números favorables (saco el mismo número)
- Luego tengo 5/6 de números favorables (saco otro número)
- Luego tengo 1/6 números favorables (saco el mismo número)
- Hay 10 maneras posibles de ordenarlo ($P'_{3,2} = 10$).

Queda $1/6 \cdot 1/6 \cdot 5/6 \cdot 1/6 \cdot 10 = 0,0385802469$
 $\Rightarrow P(\text{full}) = 0,0385802469$

Póker

4 números iguales y uno distinto
 Ejemplo: 5 5 5 3 5

Por cálculo combinatorio:

Números posibles	Formas de ordenar
vamos a usar 2 números de un total de 6, e importa cuál número será usado para el grupo de 4 dados, y cuál para el dado distinto (es decir, importa el orden)	tenemos para ordenar 4 elementos indistinguibles entre sí y otro elemento distinguible de ellos
$V_{6,2}$	$P'_{4,1}$

Queda $V_{6,2} \cdot P'_{4,1} = 150$ resultados en un total de $V'_{6,5}$ resultados posibles.
 $\Rightarrow P(\text{póker}) = 0,0192901235$

Multiplicando probabilidades:

Voy a sacar $a a a a b$ y luego lo voy a desordenar.

- Primero tengo 6/6 de números favorables (saco un número cualquiera)
- Luego tengo 1/6 de números favorables (saco el mismo número)
- Luego tengo 1/6 de números favorables (saco el mismo número)
- Luego tengo 1/6 de números favorables (saco el mismo número)
- Luego tengo 5/6 de números favorables (saco otro número)
- Hay 5 maneras posibles de ordenarlo ($P'_{4,1} = 5$).

Queda $1/6 \cdot 1/6 \cdot 1/6 \cdot 5/6 \cdot 5 = 0,0192901235$

=> $P(\text{póker}) = 0,0192901235$

Generala

5 números iguales

Ejemplo: 2 2 2 2 2

Por cálculo combinatorio:

Números posibles	Formas de ordenar
vamos a usar 1 número de un total de 6	no hay forma de desordenar, debido a que todos los dados son indistinguibles entre sí
$C_{6,1}$	1

Queda $C_{6,1} = 6$ resultados en un total de $V'_{6,5}$ resultados posibles.

=> $P(\text{generala}) = 0,0007716049$

Multiplicando probabilidades:

Voy a sacar $a a a a a$

- Primero tengo 6/6 de números favorables (saco un número cualquiera)
- Luego tengo 1/6 de números favorables (saco el mismo número)
- Luego tengo 1/6 de números favorables (saco el mismo número)
- Luego tengo 1/6 números favorables (saco el mismo número)
- Luego tengo 1/6 números favorables (saco el mismo número)

Queda $1/6 \cdot 1/6 \cdot 1/6 \cdot 1/6 = 0,0007716049$

=> $P(\text{póker}) = 0,0007716049$

Los juegos del Póker

Hay $V_{52,5} = 311875200$ manos posibles de póker. Calcularemos la probabilidad de sacar cada juego como la cantidad de formas posibles de sacar dicho juego dividido el total de manos posibles (definición de probabilidad de Laplace).

Par

2 cartas del mismo número, y las demás de otros números.

Ejemplo: 7 7 K 2 4

Por cálculo combinatorio:

Números posibles	Palos para el par	Palos para un solo	Palos para un solo	Palos para un solo	Formas de ordenar
vamos a usar 4 números de un total de 13	las cartas van a ser de 2 de 4 palos posibles	1 palo de un total de 4	1 palo de un total de 4	1 palo de un total de 4	tenemos 2 elementos indistinguibles entre sí y otros 3 elementos indistinguibles entre sí.
$V_{13,4}$	$V_{4,2}$	$V_{4,1}$	$V_{4,1}$	$V_{4,1}$	$P'_{2,3}$

Queda $V_{13,4} \cdot V_{4,2} \cdot V_{4,1} \cdot V_{4,1} \cdot V_{4,1} \cdot P'_{2,3} = 131788800$ pares posibles entre $V_{52,5}$ manos posibles

$$\Rightarrow P(\text{par}) = 0,422569028$$

Multiplicando probabilidades:

Voy a sacar $a a b c d$ y luego lo voy a desordenar.

- Primero tengo 52/52 cartas favorables (saco una carta cualquiera)
- Luego tengo 3/51 cartas favorables (las que me quedan del mismo número)
- Luego tengo 48/50 cartas favorables (para sacar otro número)
- Luego tengo 44/49 cartas favorables (para sacar otro número)
- Luego tengo 40/48 cartas favorables (para sacar otro número)
- Hay 10 maneras posibles de ordenarlo ($P'_{2,3} = 10$).

Queda $3/51 \cdot 48/50 \cdot 44/49 \cdot 40/48 \cdot 10 = 0,422569028$

$$\Rightarrow P(\text{par}) = 0,422569028$$

Par Doble

2 pares de cartas con el mismo número, y la 5ta de un 3er número.

Ejemplo: 7 7 K K 4

Por cálculo combinatorio:

Números posibles	Palos para el 1 ^{er} par	Palos para el 2 ^{do} par	Palos para el solo	Formas de ordenar
vamos a usar 3 números de un total de 13	las cartas van a ser de 2 de 4 palos posibles	las cartas van a ser de 2 de 4 palos posibles	1 palo de un total de 4 posibles	tenemos 2 elementos indistinguibles entre sí, otros 2 elementos indistinguibles entre sí, y un 5to elemento.
$V_{13,3}$	$V_{4,2}$	$V_{4,2}$	$V_{4,1}$	$P'_{2,2,1} / 2$ (*)

(*) Estamos dividiendo por 2 porque los dos pares son indistinguibles entre sí.

Queda $V_{13,3} \cdot V_{4,2} \cdot V_{4,2} \cdot V_{4,1} \cdot P'_{2,2,1} / 2 = 14826240$ manos con par doble posibles

entre $V_{52,5}$ manos posibles
 $\Rightarrow P(\text{par doble}) = 0,047539016$

Multiplicando probabilidades:

Voy a sacar $a a b b c$ y luego lo voy a desordenar.

- Primero tengo 52/52 cartas favorables (saco una carta cualquiera)
- Luego tengo 3/51 cartas favorables (las que me quedan del mismo número)
- Luego tengo 48/50 cartas favorables (para sacar otro número)
- Luego tengo 3/49 cartas favorables (las que me quedan del mismo número)
- Luego tengo 44/48 cartas favorables (para sacar otro número)
- Hay 15 maneras posibles de ordenarlo (no olvidar que los 2 pares son indistinguibles entre sí, es decir, $a a b b c$ y $b b a a c$ son lo mismo. $P'_{2,2,1} / 2 = 30/2 = 15$).

Queda $3/51 \cdot 48/50 \cdot 3/49 \cdot 44/48 \cdot 15 = 0,0475390156$
 $\Rightarrow P(\text{par doble}) = 0,0475390156$

Pierna

3 cartas del mismo número, y otras 2 con otros 2 números.
 Ejemplo: 5 9 9 A 9

Por cálculo combinatorio:

Números posibles	Palos para el trío	Palos para el 1 ^{er} solo	Palos para el 2 ^{do} solo	Formas de ordenar
vamos a usar 3 números de un total de 13	las cartas van a ser de 3 de 4 palos posibles	1 de 4 palos posibles	1 de 4 palos posibles	tenemos 3 elementos indistinguibles entre sí, y otros 2 elementos distinguibles sí
$V_{13,3}$	$V_{4,3}$	$V_{4,1}$	$V_{4,1}$	$P'_{3,2}$

Queda $V_{13,3} \cdot V_{4,3} \cdot V_{4,1} \cdot V_{4,1} \cdot P'_{3,1,1} = 6589440$ manos con pierna posibles entre $V_{52,5}$ manos posibles
 $\Rightarrow P(\text{par doble}) = 0,047539016$

Multiplicando probabilidades:

Voy a sacar $a a a b c$ y luego lo voy a desordenar.

- Primero tengo 52/52 cartas favorables (saco una carta cualquiera)
- Luego tengo 3/51 cartas favorables (las que me quedan del mismo número)
- Luego tengo 2/50 cartas favorables (las que me quedan del mismo número)
- Luego tengo 48/49 cartas favorables (para sacar otro número)
- Luego tengo 44/48 cartas favorables (para sacar otro número)
- Hay 10 maneras posibles de ordenarlo ($P'_{3,2} = 10$).

Queda $3/51 \cdot 2/50 \cdot 48/49 \cdot 44/48 \cdot 10 = 0,0211284514$
 $\Rightarrow P(\text{pierna}) = 0,0211284514$

Escalera

5 cartas con números consecutivos, considerando también el caso 10 J Q K A

Ejemplo: 10 9 7 J 8

Por cálculo combinatorio:

Escaleras posibles	Palos para cada carta	Formas de ordenar
Vamos a tomar 1 entre 10 escaleras posibles	cada una de las 5 cartas será de 1 entre 4 palos posibles.	son las formas de ordenar 5 cartas distinguibles entre sí
$V_{10,1}$	$V'_{4,5}$	P_5

Queda $V_{10,1} \cdot V'_{4,5} \cdot P_5 = 1228800$ manos con escalera posibles entre $V_{52,5}$ manos posibles

=> $P(\text{escalera}) = 0,003940038$

Multiplicando probabilidades:

En este caso la resolución mediante multiplicación de probabilidades puede tornarse muy complejo. Apreciamos entonces la ventaja de poder contar el con cálculo combinatorio.

Color

Las 5 cartas del mismo palo.

Por cálculo combinatorio:

Números posibles	Palos para el solo	Formas de ordenar
vamos a usar 5 números de un total de 13	las cartas van a ser de 1 de 4 palos posibles	tenemos para ordenar 5 elementos distinguibles
$V_{13,5}$	$V_{4,1}$	P_5

Queda $V_{13,5} \cdot V_{4,1} \cdot P_5 = 617760$ colores posibles entre $V_{52,5}$ manos posibles

=> $P(\text{color}) = 0,001980792$

Multiplicando probabilidades:

Voy a sacar una carta cualquiera y luego pediré que las 4 siguientes sean del mismo palo.

- Primero tengo 52/52 cartas favorables (saco una carta cualquiera)
- Luego tengo 12/51 cartas favorables (las que me quedan del mismo palo)
- Luego tengo 11/50 cartas favorables (las que me quedan del mismo palo)
- Luego tengo 10/49 cartas favorables (las que me quedan del mismo palo)
- Luego tengo 9/48 cartas favorables (las que me quedan del mismo palo)

Queda $12/51 \cdot 11/50 \cdot 10/49 \cdot 9/48 = 0,001980792$

=> $P(\text{color}) = 0,001980792$

Full

3 cartas del mismo número, y otras 2 del mismo número.

Ejemplo: 5 5 5 J J.

Por cálculo combinatorio:

Números posibles	Palos para la pierna	Palos para el par	Formas de ordenar
vamos a usar 2 números de un total de 13	la pierna va a usar, de un número, 3 palos de un total de 4	el par va a usar, de un número, 2 palos de un total de 4	tenemos para ordenar 3 elementos indistinguibles entre sí y otros 2 elementos indistinguibles entre sí.
$V_{13,2}$	$V_{4,3}$	$V_{4,2}$	$P'_{3,2}$

Queda $V_{13,2} \cdot V_{4,3} \cdot V_{4,2} \cdot P'_{3,2} = 449280$ fulls posibles entre $V_{52,5}$ manos posibles
 $\Rightarrow P(\text{full}) = 0,001440576$

Multiplicando probabilidades:

Voy a sacar $a a a b b$ y luego lo voy a desordenar.

- Primero tengo 52/52 cartas favorables (saco una carta cualquiera)
- Luego tengo 3/51 cartas favorables (las que me quedan del mismo número)
- Luego tengo 2/50 cartas favorables (las que me quedan del mismo número)
- Luego tengo 48/49 cartas favorables (para sacar otro número)
- Luego tengo 3/48 cartas favorables (las que me quedan del segundo número)
- Hay 10 maneras posibles de ordenarlo ($P'_{3,2} = 10$).

Queda $3/51 \cdot 2/50 \cdot 48/49 \cdot 3/48 \cdot 10 = 0,001440576$

$\Rightarrow P(\text{full}) = 0,001440576$

Póker

4 cartas del mismo número.

Ejemplo: A A A A 7.

Por cálculo combinatorio:

Números posibles	Palos para el solo	Formas de ordenar
vamos a usar 2 números de un total de 13	el que está solo va a usar 1 palo de un total de 4	tenemos para ordenar 4 elementos indistinguibles entre sí y un 1 otro elemento.
$V_{13,2}$	$V_{4,1}$	$P'_{4,1}$

Queda $V_{13,2} \cdot V_{4,1} \cdot P'_{4,1} = 74880$ pokers posibles entre $V_{52,5}$ manos posibles
 $\Rightarrow P(\text{póker}) = 0,000240096$

Multiplicando probabilidades:

Voy a sacar $a a a a b$ y luego lo voy a desordenar.

- Primero tengo 52/52 cartas favorables (saco una carta cualquiera)
- Luego tengo 3/51 cartas favorables (las que me quedan del mismo número)
- Luego tengo 2/50 cartas favorables (las que me quedan del mismo número)

- Luego tengo 1/49 carta favorable (la que me queda del mismo número)
 - Hay 5 maneras posibles de ordenarlo ($P'_{4,1} = 5$).
- Queda $3/51 \cdot 2/50 \cdot 1/49 \cdot 5 = 0,000240096$
 $\Rightarrow P(\text{póker}) = 0,000240096$

Escalera Real

Es tener escalera y color al mismo tiempo

Por cálculo combinatorio:

Escaleras posibles	Palos para cada carta	Formas de ordenar
vamos a tomar 1 entre 10 escaleras posibles	la escalera será de 1 de 4 palos posibles	son las formas de ordenar 5 cartas distinguibles entre sí
$V_{10,1}$	$V_{4,1}$	P_5

Queda $V_{10,1} \cdot V_{4,1} \cdot P_5 = 4800$ manos con escalera real posibles entre $V_{52,5}$ manos posibles

$\Rightarrow P(\text{escalera real}) = 0,0000153908$

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 10 de junio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 6 de junio de 2004

APÉNDICE B

Otros problemas de probabilidad

En este capítulo estudiaremos la forma de resolver algunos problemas que no tienen que ver con ninguna distribución en particular sino que podrían ser aplicados a distintas distribuciones o combinaciones de ellas.

Este capítulo es, en consecuencia, netamente práctico.

Sumar cantidades desconocidas de probabilidades

"En promedio, el 25% de las piezas son defectuosas. En un lote de 10 piezas, calcule la máxima cantidad de piezas defectuosas que se podrá garantizar, tal que haya 90% de que se cumpla la predicción"

Por lo general en todo problema simple de probabilidad los elementos involucrados son:

- una distribución
- los parámetros de la distribución
- una variable que está distribuida según esa distribución
- una condición sobre la variable
- la probabilidad de que se cumpla esa condición

Por ejemplo, si tomamos el enunciado que figura más arriba, el problema es:

X: Binomial($n = 10$; $p = 0,25$)

$P(X \leq 4) = 0,92187$

Y los elementos involucrados son:

- distribución: binomial
- parámetros: $n = 10$; $p = 0,25$
- variable: X
- condición: $X \leq 4$
- probabilidad: 0,92187

Generalmente los problemas consisten en tener los 4 primeros ítems y averiguar la probabilidad. Pero a veces la probabilidad se conoce, porque se ha calculado experimentalmente, porque se quiere tener una probabilidad en particular, etc. Y entonces el problema consiste en encontrar alguno de los otros 4 elementos. El problema que vamos a estudiar ahora es el caso en el cual hay que averiguar la condición.

En el ejemplo que vimos, nos están pidiendo que encontremos un valor tal que haya una probabilidad 0,9 de que la X resulte menor o igual que ese valor.

Es decir, nos piden encontrar m tal que:

$$P(X \leq m) = 0,9$$

Como la distribución es binomial, sabemos que $P(X = x)$ tendrá valores no nulos solamente para $x \in [0 ; n]$. Y podemos asumir que $0 \leq m \leq n$.

Entonces:

$$P(X \leq m) = \sum_{i=0}^m P(X = i) = 0,9$$

Como podemos ver, el problema consistirá en ir acumulando $P(X = 0) + \dots + P(X = m)$ hasta llegar a 0,9.

Usando la fórmula de la binomial obtenemos:

$$P(X = 0) = 0,05631$$

$$P(X = 1) = 0,18771 \quad \Rightarrow \quad P(X \leq 1) = 0,24403$$

$$P(X = 2) = 0,28157 \quad \Rightarrow \quad P(X \leq 2) = 0,52559$$

$$P(X = 3) = 0,25028 \quad \Rightarrow \quad P(X \leq 3) = 0,77588$$

$$P(X = 4) = 0,14600 \quad \Rightarrow \quad P(X \leq 4) = 0,92187$$

Vemos que si tomamos $m = 4$ nos pasamos de 0,9. ¿Entonces tomamos 3 ó tomamos 4?

La clave está en el enunciado. Nos piden un m que **garantice** que la probabilidad sea 0,9. Intuitivamente se entiende que $m = 4$ garantiza esa probabilidad, porque da mayor, y que $m = 3$ no la garantiza. Entonces se elige el 4.

Tenemos que $P(X \leq 4) \geq 0,9$.

De hecho en vez de tener que analizar para dónde vamos a redondear, podríamos cambiar nuestro planteo original $P(X \leq m) = 0,9$ por $P(X \leq m) \geq 0,9$ puesto que se nos pide **garantizar** la probabilidad.

En conclusión, si informamos que la respuesta es 4, estamos proveyendo el menor número entero tal que la probabilidad de que en un lote cualquiera la cantidad de piezas defectuosas sea menor o igual a él es igual o mayor a 0,9.

Por último, notemos que el mismo problema podría haber sido enunciado de forma

ligeramente distinta:

"En promedio, el 25% de las piezas son defectuosas. Si se arman lotes de 10 piezas, calcule la máxima cantidad de piezas defectuosas que se podrá garantizar en el 90% de los lotes"

El problema es exactamente el mismo.

Ahora veremos otros ejemplos:

Problemas típicos

1) Cierta artefacto de pirotecnia está diseñado para producir 100 explosiones. Sin embargo, el 1.2% de las explosiones falla. Calcule:

- a) **¿Cuánto es el máximo de explosiones que fallarán en un artefacto, con un 95% de confianza?**
- b) **¿Cuántas explosiones podrá Ud. garantizarle al cliente, con un 95% de confianza?**

Resolución:

a) ¿Qué significa "95% de confianza"? Significa que si informamos que la cantidad máxima de fallos es m , entonces tendremos probabilidad 0,95 de que cada artefacto tenga m o menos fallas.

Si X es la cantidad de fallas, entonces estamos buscando m tal que:

$$P(X \leq m) \geq 0,95$$

Entonces el problema se reduce a ir acumulando $P(X = 0) + \dots + P(X = m)$ hasta encontrar m_0 tal que la sumatoria llegue a 0,95. Es decir, buscamos m tal que:

$$\sum_{i=0}^m P(X = i) \geq 0,95$$

¿Y cómo calculamos $P(X = i)$? Aunque el lector habrá adivinado que la distribución es binomial, es importante destacar que todavía no hemos usado el hecho de que la distribución es binomial. En este tipo de problemas, conocer cuál distribución es la que vamos a emplear por lo general solo se vuelve necesario a la hora de hacer los cálculos. Por esa razón, ejercicios de este tipo no fueron incluidos en la sección de problemas típicos de cada sección.

Se puede ver que la cantidad de fallas en 100 explosiones está distribuida binomialmente con $n = 100$ y $p = 0,012$.

$$\Rightarrow X: \text{Bi}(n = 100 ; p = 0,012)$$

Usando la fórmula de la binomial obtenemos:

$$P(X = 0) = 0,29902$$

$$P(X = 1) = 0,36318 \quad \Rightarrow \quad P(X \leq 1) = 0,66219$$

$$P(X = 2) = 0,21835 \quad \Rightarrow \quad P(X \leq 2) = 0,88054$$

$$P(X = 3) = 0,08663 \quad \Rightarrow \quad P(X \leq 3) = 0,96717$$

Esto implica que si informamos que el máximo es 3, la probabilidad de que un artefacto tenga 3 o menos fallas es 0,96717 lo cual supera el 0,95 que nos pidieron.

b) Este problema es el mismo que en el caso "a", pero planteado de diferente forma. Comenzaremos por hacer el mismo análisis que en "a":

"95% de confianza" significa que si informamos que la cantidad de explosiones es c , tendremos probabilidad 0,95 de que cada artefacto tenga al menos c explosiones. Entonces si X es la cantidad de explosiones, entonces $X \sim \text{Bi}(n = 100 ; p = 0,988)$.

Y estamos buscando c tal que $P(X \geq c) \geq 0,95$.

Ahora debemos resolver eso. Tenemos 2 formas:

Forma 1:

$P(X \geq c)$ será una sumatoria como en "a", pero con la diferencia de que como en este caso estamos plantando X mayor o igual a algo, en vez de menor o igual como en "a", la sumatoria nos dará en sentido decreciente. Es decir, estaremos buscando c tal que:

$$P(X = 100) + P(X = 99) + \dots + P(X = c) \geq 0,95$$

es decir:

$$\sum_{i=c}^{100} P(X = i) \geq 0,95$$

Usando la fórmula de la binomial obtenemos:

$$P(X = 100) = 0,29902$$

$$P(X = 99) = 0,36318 \Rightarrow P(X \geq 99) = 0,66219$$

$$P(X = 98) = 0,21835 \Rightarrow P(X \geq 98) = 0,88054$$

$$P(X = 97) = 0,08663 \Rightarrow P(X \geq 97) = 0,96717$$

Con lo cual si le garantizamos al cliente que el artefacto producirá 97 explosiones, el 96,7% de los artefactos efectivamente cumplirán con dicha especificación.

Forma 2:

$$P(X \geq c) \geq 0,95$$

$$1 - P(X < c) \geq 0,95$$

$$P(X < c) \leq 0,05$$

$$P(X \leq c-1) \leq 0,05$$

$$\sum_{i=0}^{c-1} P(X = i) \leq 0,05$$

Usando la fórmula de la binomial obtenemos:

$$P(X = 0) \approx 0$$

$$P(X = 1) \approx 0 \Rightarrow P(X \leq 1) \approx 0$$

$$P(X = 2) \approx 0 \Rightarrow P(X \leq 2) \approx 0$$

...

$$\begin{aligned}P(X = 91) &\approx 0 &\Rightarrow P(X \leq 91) &\approx 0 \\P(X = 92) &= 0,00003 &\Rightarrow P(X \leq 92) &= 0,00003 \\P(X = 93) &= 0,00019 &\Rightarrow P(X \leq 93) &= 0,00022 \\P(X = 94) &= 0,00114 &\Rightarrow P(X \leq 94) &= 0,00136 \\P(X = 95) &= 0,00595 &\Rightarrow P(X \leq 95) &= 0,00731 \\P(X = 96) &= 0,02552 &\Rightarrow P(X \leq 96) &= 0,03283 \\P(X = 97) &= 0,08663 &\Rightarrow P(X \leq 97) &= 0,11946\end{aligned}$$

Con lo cual llegamos al mismo resultado $c = 97$, pero hicimos una engorrosa sumatoria de 97 términos. Esto nos muestra que cuando nos encontremos con planteos que nos lleven a resoluciones así, es conveniente replantear el problema para poder resolverlo como en la forma 1.

2) Una máquina se rompe en promedio 3 veces por año. ¿Cuál es la cantidad máxima de fallas en un año que se le pueden prometer a un cliente con un 95% de confianza?

Resolución:

95% de confianza significa en este caso que estamos buscando cuál es la cantidad máxima de fallas que podemos prometer, con probabilidad 0,95 de que nuestra promesa se cumpla.

Si X es la cantidad de fallas en un año, entonces buscamos c tal que:

$$P(X \leq c) \geq 0,95$$

Es decir:

$$P(X = 0) + P(X = 1) + \dots + P(X = c) \geq 0,95$$

Si esa variable X está distribuida según poisson, su media es $\mu = 3 \text{ veces/año} \cdot 1 \text{ año} = 3$ veces = 3

Usando la fórmula de poisson obtenemos:

$$\begin{aligned}P(X = 0) &= 0,04979 \\P(X = 1) &= 0,14936 &\Rightarrow P(X \leq 1) &= 0,19915 \\P(X = 2) &= 0,22404 &\Rightarrow P(X \leq 2) &= 0,42319 \\P(X = 3) &= 0,22404 &\Rightarrow P(X \leq 3) &= 0,64723 \\P(X = 4) &= 0,16803 &\Rightarrow P(X \leq 4) &= 0,81526 \\P(X = 5) &= 0,10082 &\Rightarrow P(X \leq 5) &= 0,91608 \\P(X = 6) &= 0,05041 &\Rightarrow P(X \leq 6) &= 0,96649\end{aligned}$$

Con lo cual si le informamos al cliente que la máquina presentará menos de 6 fallas en un año, habrá probabilidad 0,96649 de que la predicción se cumpla.

Considerar varias distribuciones al mismo tiempo

Hay problemas en los que debemos considerar más de una distribución al mismo tiempo, por ejemplo:

- cuando hay que sumar una determinada cantidad de variables, y la cantidad a sumar no es una constante sino que está dada por otra variable.
- cuando una distribución tiene un parámetro que no es una constante sino que se calcula mediante otra distribución.

Estos problemas son muy útiles para el estudiante de probabilidad y estadística debido a su carácter integrador: para resolverlos es necesario conocer muchos temas distintos y desarrollar la capacidad de relacionarlos.

Para entender los ejemplos que se dan a continuación, es necesario comprender bien el capítulo 3, así como conocer bien, según corresponda en cada ejemplo, las distribuciones estudiadas en los capítulos 4, 5, 6 y 7.

Problemas típicos

1) Juan usa su auto para ir y volver del trabajo. En el viaje de ida y vuelta, la cantidad de litros de combustible consumida por el auto es una variable exponencial negativa con $\lambda = 0.4$. Si la cantidad de días por semana que trabaja está distribuida binomialmente con $n = 5$ y $p = 0,9$. ¿Cuál es la probabilidad de que en una semana su auto consuma menos de 13 litros?

Resolución

Este ejercicio suma una cantidad de exponenciales negativas dada por una binomial. El consumo de cada día en que se usa el auto es una $X_i: \text{ExpNeg}(0.4)$. La cantidad de días que trabaja en una semana es $Y: \text{Bi}(5; 0,9)$. El consumo en una semana es:

$$Z = \sum_{i=1}^Y X_i$$

Z será entonces una suma de exponenciales negativas. Luego Z tiene una distribución gamma con parámetro λ igual al de las X_i y parámetro $k = Y$. Pero como Y no es una constante sino una variable aleatoria, la expresión de la función de densidad de la gamma no será en realidad f_Z sino $f_{Z/Y}$. Es decir:

$$f_{Z/Y} = \begin{cases} \frac{\lambda (\lambda z)^{y-1} e^{-\lambda z}}{(y-1)!} & z \geq 0 \\ 0 & z < 0 \end{cases}$$

Tenemos que $f_{Z/Y}$ es simplemente f_Z , pero haciendo eco del hecho de que uno de los parámetros, en este caso k , no es una constante sino una variable aleatoria. De

hecho siempre se puede escribir la función de densidad de una variable aleatoria como la distribución condicional a los parámetros. Es decir, si Z es una variable aleatoria gamma, escribir f_Z y $f_{Z/\lambda,k}$ es lo mismo. Sucede que como los parámetros por lo general son constantes, y no variables aleatorias como en este caso, por simplicidad las funciones de densidad no se escriben como condicionales.

Vamos a usar la ya estudiada técnica de hacer intersección con el espacio muestral

(probabilidad total), para hacer aparecer todos los valores de Y :

$$P(Z < 13) = P(Z < 13 \cap E) = P(Z < 13 \cap (Y = 0 \cup \dots \cup Y = 5)) = P(Z < 13 \cap Y = 0 \cup \dots \cup Z < 13 \cap Y = 5) = \\ = P(Z < 13 \cap Y = 0) + \dots + P(Z < 13 \cap Y = 5) = P\left(\frac{Z < 13}{Y = 0}\right)P(Y = 0) + \dots + P\left(\frac{Z < 13}{Y = 5}\right)P(Y = 5)$$

Nos quedó una sumatoria de productos entre las condicionales, y las probabilidades de Y .

Dichas condicionales son simples probabilidades de la gamma:

$$P\left(\frac{Z < 13}{Y = y}\right) \text{ donde } y \text{ es el valor que asume } k \text{ en cada una de ellas.}$$

Según estudiamos, podemos calcular las probabilidades de una gamma con una Poisson, usando:

$$\int_0^{z_0} f_Z(z) dz = 1 - \sum_{i=0} P(W = i)$$

donde W es una Poisson con parámetro $\mu = \lambda z_0$.

Luego:

$$P\left(\frac{Z < 13}{Y = y}\right) = P\left(\frac{Z < 13}{k = y}\right) = \int_0^{13} f_Z(z) dz = 1 - \sum_{i=0} P(W = i) \\ P(W = w) = \frac{e^{-5,2} 5,2^w}{w!}$$

donde $W: \text{Pois}(\mu = 13\lambda = 5,2)$, con lo cual

Calculamos:

$$P\left(\frac{Z < 13}{Y = 0}\right) = 1$$

$$P\left(\frac{Z < 13}{Y = 1}\right) = 1 - P(W = 0) = 0,99448$$

$$P\left(\frac{Z < 13}{Y = 2}\right) = 1 - P(W = 0) - P(W = 1) = 0,96580$$

$$P\left(\frac{Z < 13}{Y = 3}\right) = 1 - P(W = 0) - P(W = 1) - P(W = 2) = 0,89121$$

$$P\left(\frac{Z < 13}{Y = 4}\right) = 1 - P(W = 0) - P(W = 1) - P(W = 2) - P(W = 3) = 0,76193$$

$$P\left(\frac{Z < 13}{Y = 5}\right) = 1 - P(W = 0) - P(W = 1) - P(W = 2) - P(W = 3) - P(W = 4) = 0,59387$$

Las probabilidades $P(Y = y)$ que van multiplicadas por estas condicionales se obtienen usando la fórmula para las probabilidades binomiales. Calculándolas, multiplicándolas por las condicionales, y sumando todo, queda: 0,6739.

2) Se hace un viaje de 400km. Por diversos problemas, deben hacerse diversas detenciones, a razón de una detención cada 90 km, en promedio. El tiempo perdido en cada detención es una variable aleatoria $N(15 ; 2)$ en

minutos.

- a) Calcule el tiempo esperado que se perderá en detenciones.
- b) Halle la fórmula que usaría para calcular la probabilidad de que haya habido 5 detenciones si se perdió menos de una hora en detenciones.

Resolución

Este ejercicio suma una cantidad de normales dada por una Poisson. La demora en cada detención es una $X_i: N(15; 2)$. La cantidad de detenciones es $Y: \text{Pois}(\mu)$ donde $\mu = \lambda_t = 1/90 \cdot 400 = 4,44$. El total de demoras en detenciones es:

$Z = \sum_{i=1}^Y X_i$, con lo cual Z es una normal $N(15Y; 2\sqrt{Y})$. Al igual que vimos en el ejercicio anterior, lo que tenemos en realidad no es la distribución de Z , sino la distribución de Z dado Y .

a) $E(Z) = E(15Y) = 15 E(Y) = 15 \cdot 4,44 = 66,67$.

$$P(Y=5/Z < 60) = \frac{P(Z < 60/Y=5) P(Y=5)}{P(Z < 60)}$$

b)

Y luego:

$$P(Z < 60/Y=5) = F_{Z/Y=5}(60) = \Phi\left(\frac{60 - 75}{2\sqrt{5}}\right) = 0,0004$$

$$P(Y=5) = \frac{e^{-4,44} 4,44^5}{5!} = 0,1697$$

$P(Z < 60)$ queda, usando probabilidad total como ya se vio en el problema anterior:

$$\sum_{y=0} P(Z < 60/Y=y) P(Y=y) = \sum_{y=0} \Phi\left(\frac{60 - 15y}{2\sqrt{y}}\right) \frac{e^{-4,44} 4,44^y}{y!}$$

En conclusión la fórmula sería:

$$P(Y=5/Z < 60) = \frac{0,0004 \cdot 0,1697}{\sum_{y=0} \Phi\left(\frac{60 - 15y}{2\sqrt{y}}\right) \frac{e^{-4,44} 4,44^y}{y!}}$$

y con esa fórmula se pueden sumar términos hasta que se hagan despreciables y calcular la probabilidad buscada.

3) Un determinado hilo de cobre se vende en bobinas de 15 metros. Dicho hilo presenta defectos de fabricación a razón de un defecto cada 60 metros. Un cliente está interesado en comprar 1000 bobinas, pero para asegurarse de que sean de buena calidad, revisa el 1% de las bobinas, y cancela la compra si encuentra 2 ó más bobinas con defectos. ¿Cuál es la probabilidad de que el cliente cancele la compra?

Resolución

En este ejercicio se usa una Poisson para calcular el parámetro p de una binomial.

El 1% de las 1000 bobinas son 10 bobinas. El cliente cancela la compra si encuentra 2 ó más bobinas con defectos en ese lote de 10. Si $X:Bi(n=10;p)$, donde p es la probabilidad de que una bobina tenga defectos, entonces:

$$P(\text{el cliente cancela la compra}) = P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$$

Necesitamos calcular p .

Si Y es la cantidad de defectos en una bobina, entonces $Y:Pois(\mu)$ donde:

$$\mu = 1/60 \cdot 15 = 1/4$$

La probabilidad de que una bobina tenga defectos es $P(Y \geq 0) = 1 - P(Y = 0) = 1 - 0,68729 = 0,2212 = p$.

Luego la probabilidad de que el cliente cancele la compra es:

$$1 - P(X = 0) - P(X = 1) = 1 - 0,08208 - 0,23314 = 0,68477$$

4) Juan tiene asignadas para dormir 8 horas de cada noche. Sin embargo, durante dichas horas, el teléfono suena en promedio una vez cada tres horas. Calcule:

- a) La probabilidad de que en una semana pueda dormir más de 2 noches sin que lo despierten.
- b) La probabilidad de que recién a la cuarta noche que lo intente pueda dormir sin que lo despierten.

Resolución:

En este ejercicio se usa una Poisson para calcular el parámetro p de una binomial y una geométrica.

a) Si X es la cantidad de noches en las que no lo despiertan, $X:Bi(n = 7 ; p)$ donde p es la probabilidad de que en una noche no lo despierten.

La probabilidad de que en una noche no lo despierten es la probabilidad de que en 8 horas haya cero llamadas, es decir, si Y es la cantidad de llamadas, $Y:Pois(\mu)$ donde $\mu = 1/3 \cdot 8 = 8/3$ entonces buscamos:

$$P(Y = 0) = 0,06948 = p$$

Luego la probabilidad de que pueda dormir en una semana más de dos noches sin que lo despierten es:

$$P(X > 2) = 1 - P(X = 0) - P(X = 1) - P(X = 2) = 0,00949.$$

b) Si Z es la primera noche en la que logra dormir sin que lo despierten, entonces $Z:Geom(p)$.

$$\text{Luego } P(Z = 4) = 0,05598.$$

5) Las varillas deben medir $20 \pm 0,1$ cm. La longitud de las varillas que fabrica la máquina es una variable $N(20 ; 0,05)$. ¿Cuál es la probabilidad de que en un lote de 10 varillas haya alguna con longitud fuera del rango permitido?

Resolución

En este ejercicio se usa una normal para encontrar el parámetro p de una binomial. Si X es la longitud de una varilla, $X:N(20 ; 0,05)$, y la probabilidad de que una varilla se encuentre fuera del rango permitido vale:

$$1 - P(19,9 < X < 20,1)$$

$$1 - P(19,9 < X < 20,1) = 1 - (F_X(20,1) - F_X(19,9)) = 1 - F_X(20,1) + F_X(19,9) = 1 - \Phi\left(\frac{20,1 - 20}{0,05}\right) + \Phi\left(\frac{19,9 - 20}{0,05}\right) = 1 - 0,97725 + 0,02275 = 0,0455$$

Luego ese valor es el p de $Y:Bi(n = 10 ; p)$ que representa la cantidad de varillas no aceptables en el lote de 10. Se pide:

$$P(Y \geq 0) = 1 - P(Y = 0) = 0,37229.$$

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 6 de junio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 17 de junio de 2004

APÉNDICE C

Simulación

Si una población sigue una determinada distribución, entonces al tomar una muestra de n valores de esa población, los valores obtenidos cumplirán con 2 características:

- 1) todos ellos serán valores posibles (es decir, son valores que tienen probabilidad no-nula en la distribución de la población)
- 2) las proporciones entre los valores cumplirán aproximadamente con la forma de la distribución.

Por ejemplo, si la población fueran las duraciones de determinados componentes producidos, y dichas duraciones siguen una distribución exponencial negativa, y se toma una muestra (es decir, se eligen n componentes al azar y se miden sus duraciones) entonces:

- 1) los valores obtenidos serán números reales positivos (la distribución exponencial negativa le asigna probabilidad no nula a los valores reales positivos)
- 2) la proporción entre los valores cumplirá aproximadamente con la forma de la distribución exponencial negativa, es decir, la cantidad de valores obtenidos cercanos al cero seguramente será mayor que la cantidad de valores obtenidos lejanos al cero, porque en la distribución exponencial negativa los valores más probables son los más cercanos al cero).

Esto se entiende porque como vimos en los capítulos anteriores, cada elemento que compone nuestra muestra no es otra cosa que una variable aleatoria cuya distribución es la de la población de la cual extraemos la muestra.

El problema de la simulación consiste en, dada una determinada distribución, generar un conjunto de valores que podrían haber venido de una población que tenga esa distribución, es decir, que sean valores posibles, y que las proporciones entre ellos reflejen la distribución de la cual deben parecer venir. La diferencia entre tomar una muestra y simular, es que en la muestra los valores se obtienen extrayendo elementos de una población, mientras que en la simulación se "inventan".

Por ejemplo, si la población de la cual queremos simular una muestra tiene la distribución $U(0;1)$ (capítulo 7) entonces la podemos simular con la función "random" de cualquier calculadora o computadora. Dicha función nos provee cada vez que la invocamos de un número al azar entre cero y uno. Notemos que obtener

un número al azar entre cero y uno, no es otra cosa que una simulación de una distribución $U(0;1)$. Luego para obtener una muestra simulada de tamaño n de una distribución $U(0;1)$ basta con utilizar n veces la función random de la calculadora o computadora.

¿Cómo hacemos si la distribución que queremos simular no es $U(0;1)$? Como dijimos antes, cada elemento de una muestra que se toma es una variable aleatoria que tiene la distribución de la población. Entonces para poder simular tenemos que encontrar la manera de crear artificialmente una variable aleatoria cuya distribución sea la de la población para la cual queremos simular una muestra.

Para esto, nos valdremos justamente de la función random. Sabemos que el valor arrojado por la función random es una variable aleatoria $X:U(0;1)$, y lo que queremos simular es una variable aleatoria cualquiera Y , cuya distribución viene dada por $f_Y(y)$. Lo que haremos será tomar un cambio de variables $Y = \Phi(X)$, de modo tal que dados los valores de X (que podemos obtener fácilmente) mediante un pequeño cálculo obtengamos los valores de Y . Entonces nuestro problema se reduce a encontrar un cambio de variables adecuado, que nos garantice que si la distribución de X es $U(0;1)$ entonces la distribución de $Y = \Phi(X)$ sea la $f_Y(y)$ que queremos simular.

Como cambio de variables, vamos a proponer la función $Y = \Phi(x) = F_Y^{-1}(x)$. Como vimos en el capítulo 2, la fórmula para encontrar la distribución de Y es:

$$f_Y(y) = \frac{f_X(\Phi^{-1}(x))}{\left| \frac{dy}{dx} \right|} = f_X(\Phi^{-1}(x)) \left| \frac{dx}{dy} \right|$$

Como se vio en el capítulo 7, si $X:U(0;1)$ entonces la función $f_X(x)$ vale:

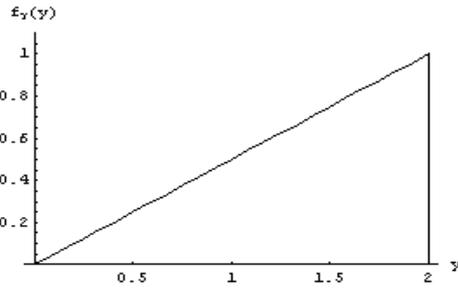
$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \forall \text{ otro } x \end{cases}$$

Luego como dentro del dominio de X , $f_X(x)$ siempre vale 1, queda:

$$f_Y(y) = \left| \frac{dx}{dy} \right|$$

Decir que $Y = F_Y^{-1}(x)$ es lo mismo que decir que $X = F_Y(y)$. Luego, dx/dy es $f_Y(y)$. Por lo tanto, hemos demostrado que si X es uniforme entre 0 y 1, y dada $f_Y(y)$ una distribución cualquiera que queremos simular, entonces si tomamos el cambio de variables $Y = \Phi(x) = F_Y^{-1}(x)$, los valores que obtendremos para Y tendrán la distribución $f_Y(y)$ que queríamos simular.

Ejemplo



Simularemos a continuación 10 valores de la siguiente distribución:

$$f_Y(y) = \begin{cases} y/2 & 0 < y < 2 \\ 0 & \forall \text{ otro } y \end{cases}$$

Vamos a necesitar $F_Y^{-1}(y)$. Construimos $F_Y(y)$:

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ y^2/4 & 0 < y < 2 \\ 1 & y > 2 \end{cases}$$

Solamente nos interesa la rama $0 < y < 2$. Si $X = F_Y(y) = y^2/4$ entonces

$$Y = \sqrt{4x} = 2\sqrt{x}$$

Usando la función random de cualquier calculadora obtenemos los valores:

0.313, 0.579, 0.168, 0.812, 0.247, 0.324, 0.759, 0.499, 0.991, 0.117

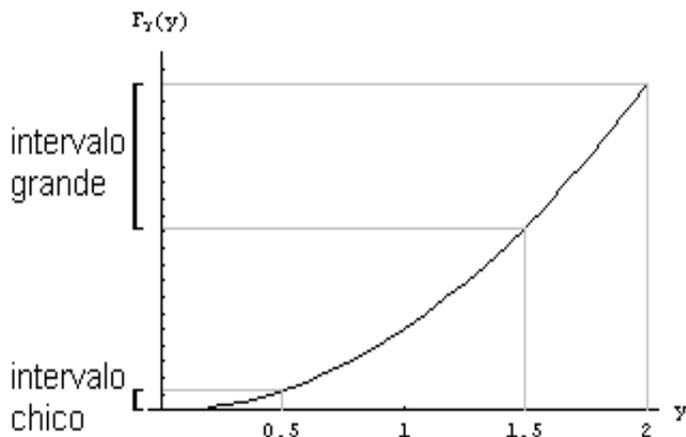
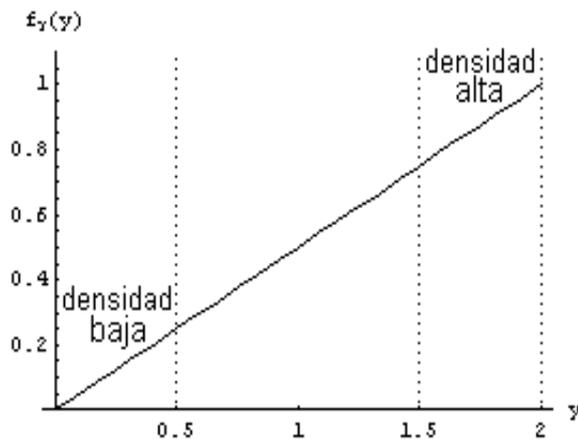
Luego aplicándole a esos valores la transformación $Y = 2\sqrt{x}$ obtenemos:

1.12, 1.52, 0.82, 1.80, 0.99, 1.14, 1.74, 1.41, 1.99, 0.68

Estos valores constituyen nuestra simulación de tamaño 10 de la variable aleatoria dada por la $f_Y(y)$ de la que partimos. Mirándolos vemos que efectivamente parecen bastante representativos de la distribución estudiada, porque predominan los valores cercanos al 2.

Resuelto el ejemplo, volveremos sobre una pregunta que quedó pendiente: ¿por qué se nos ocurrió proponer $Y = \Phi(x) = F_Y^{-1}(x)$ como solución al problema de la simulación?

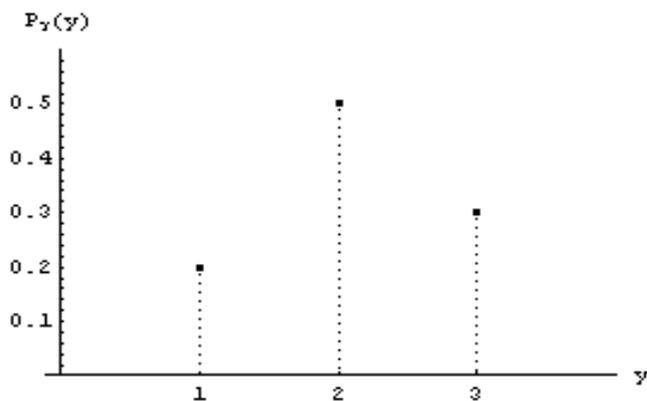
Observemos que el dominio de la función $F_Y(y)$ son los números reales, y que al ser la función de probabilidad acumulada, su imagen es el intervalo $[0;1]$. Luego la inversa $F_Y^{-1}(x)$ irá del intervalo $[0;1]$ a los reales. Más precisamente, si el número que recibe está en el intervalo $(0;1)$, $F_Y^{-1}(x)$ nos devolverá un valor posible de la variable aleatoria Y .



Ese valor tendrá la distribución deseada $f_Y(y)$. Por ejemplo, en los lugares donde $f_Y(y)$ sea alta, $F_Y(y)$ crecerá rápidamente, es decir que una pequeña porción del dominio de $F_Y(y)$ estará asociada a una gran porción de la imagen $[0;1]$. Luego una gran porción del dominio de $F_Y^{-1}(x)$ estará asociada a una pequeña porción de la imagen de $F_Y^{-1}(x)$, o sea de los valores de Y , con lo cual habrá probabilidad alta de que un número random caiga en la porción asociada a los valores correspondientes de la variable Y . Luego vemos que si en una región $f_Y(y)$ es alta, efectivamente se cumple que habrá alta probabilidad de que muchos valores simulados caigan en esa región. Comprobamos entonces que este método para simular es coherente.

Variables discretas

Para variables discretas, el método de tomar $Y = \Phi(x) = F_Y^{-1}(x)$ con X random sigue siendo válido. De hecho resulta más simple, porque en vez de encontrar la expresión de la función inversa $F_Y^{-1}(x)$ se puede directamente ver en qué región del dominio de $F_Y(x)$ cae cada valor X .



Ejemplo

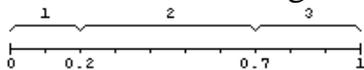
Simularemos a continuación 10 valores de la siguiente distribución:

$$P_Y(y) = \begin{cases} 0.2 & y=1 \\ 0.5 & y=2 \\ 0.3 & y=3 \\ 0 & \forall \text{ otro } y \end{cases}$$

Construyendo la función $F_Y(y)$ obtenemos:

$$F_Y(y) = \begin{cases} 0 & y < 1 \\ 0.2 & 1 \leq y < 2 \\ 0.7 & 2 \leq y < 3 \\ 1 & y \geq 3 \end{cases}$$

Podemos hacer un gráfico de este estilo:



Luego, dados los valores random, basta con ver en qué intervalo caen para saber a qué valor de Y están asociados. Si los valores random que obtenemos son: 0.057, 0.532, 0.639, 0.346, 0.588, 0.920, 0.888, 0.511, 0.841, 0.382

Entonces los valores simulados de Y serán:

1, 2, 2, 2, 2, 3, 3, 2, 3, 2

Vemos que obtuvimos 10% de 1, 60% de 2 y 30% de 3, proporciones bastante parecidas a las probabilidades respectivas 0.2, 0.5 y 0.3 de la distribución que simulamos. Cuando mayor sea el tamaño de la muestra simulada, más tenderán a parecerse las proporciones de los valores simulados a las probabilidades correspondientes.

Problemas típicos

1) Simule 10 valores de una distribución exponencial negativa con $\lambda = 2$.

Resolución

Si $Y: \text{Expneg}(\lambda = 2)$, entonces:

$$f_Y(y) = \begin{cases} 2e^{-2y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

Luego la función de distribución acumulada es:

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ 1 - e^{-2y} & y > 0 \end{cases}$$

Tenemos que $X = F_Y(y) = 1 - e^{-2y}$. Luego:

$$y = -\frac{1}{2} \ln(1 - x)$$

Obtenemos 10 valores random:

0.057, 0.532, 0.639, 0.346, 0.588, 0.920, 0.888, 0.511, 0.841, 0.382

Ahora aplicamos la transformación para obtener los valores simulados de Y:

0.029, 0.380, 0.509, 0.212, 0.443, 1.263, 1.095, 0.358, 0.919, 0.241

2) Simule 10 valores de una variable aleatoria binomial con $n = 3$ y $p = 0.8$

Resolución

Si $Y: \text{Bi}(n = 3 ; p = 0.8)$, entonces:

$$P_Y(y) = \begin{cases} 0.027 & y = 1 \\ 0.189 & y = 2 \\ 0.441 & y = 3 \\ 0.343 & y = 4 \\ 0 & \forall \text{ otro } y \end{cases}$$

Construyendo la función $F_Y(y)$ obtenemos:

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ 0.027 & 0 \leq y < 1 \\ 0.216 & 1 \leq y < 2 \\ 0.657 & 2 \leq y < 3 \\ 1 & y \geq 3 \end{cases}$$

Procediendo como en el ejemplo, asignamos los siguientes intervalos a los siguientes valores:

- $[0 ; 0.027) \rightarrow 0$
- $[0.027 ; 0.216) \rightarrow 1$
- $[0.216 ; 0.657) \rightarrow 2$
- $[0.657 ; 1) \rightarrow 3$

Luego, si los valores random fueran por ejemplo:

0.685, 0.012, 0.960, 0.833, 0.551, 0.699, 0.320, 0.227, 0.918, 0.175

Entonces los valores simulados son:

3, 0, 3, 3, 2, 2, 2, 2, 3, 1

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

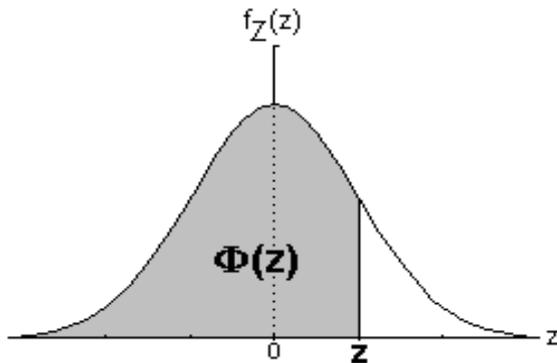
Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 17 de junio de 2004_

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004



Distribución normal estándar acumulada (Φ)

Dada $Z:N(0;1)$, $\Phi(z) = P(Z \leq z)$

Si z es negativo, se puede usar $\Phi(z) = 1 - \Phi(-z)$

Ejemplos:

- $P(Z \leq 0.38) = \Phi(0.38) = 0.64803$
- $P(Z \leq -1.52) = \Phi(-1.52) = 1 - \Phi(1.52) = 1 - 0.93574 = 0.06426$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807

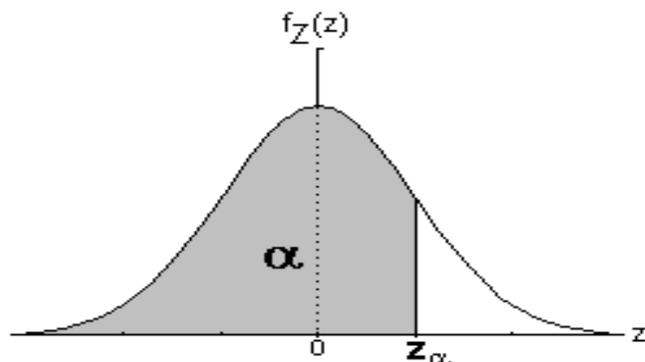
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg

<alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004



Fractiles de la normal estándar (z_α)

Dada $Z:N(0;1)$, $z_\alpha = z$ tal que $\Phi(z) = P(Z \leq z) = \alpha$

Si $\alpha < 0.5$, se puede usar $z_\alpha = -z_{1-\alpha}$

Ejemplos:

- $z_{0.617} = 0.29761$
- $z_{0.123} = -z_{0.877} = 0.16012$

	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.50	0.00000	0.00251	0.00501	0.00752	0.01003	0.01253	0.01504	0.01755	0.02005	0.02256
0.51	0.02507	0.02758	0.03008	0.03259	0.03510	0.03761	0.04012	0.04263	0.04513	0.04764
0.52	0.05015	0.05266	0.05517	0.05768	0.06019	0.06271	0.06522	0.06773	0.07024	0.07276
0.53	0.07527	0.07778	0.08030	0.08281	0.08533	0.08784	0.09036	0.09288	0.09540	0.09791
0.54	0.10043	0.10295	0.10547	0.10799	0.11052	0.11304	0.11556	0.11809	0.12061	0.12314
0.55	0.12566	0.12819	0.13072	0.13324	0.13577	0.13830	0.14084	0.14337	0.14590	0.14843
0.56	0.15097	0.15350	0.15604	0.15858	0.16112	0.16366	0.16620	0.16874	0.17129	0.17383
0.57	0.17637	0.17892	0.18147	0.18402	0.18657	0.18912	0.19167	0.19422	0.19678	0.19934
0.58	0.20189	0.20445	0.20701	0.20957	0.21214	0.21470	0.21727	0.21983	0.22240	0.22497
0.59	0.22755	0.23012	0.23269	0.23527	0.23785	0.24043	0.24301	0.24559	0.24817	0.25076

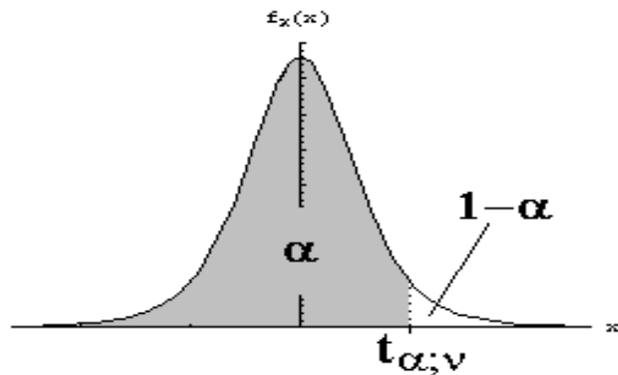
0.60	0.25335	0.25594	0.25853	0.26112	0.26371	0.26631	0.26891	0.27151	0.27411	0.27671
0.61	0.27932	0.28193	0.28454	0.28715	0.28976	0.29238	0.29499	0.29761	0.30023	0.30285
0.62	0.30548	0.30811	0.31074	0.31337	0.31600	0.31864	0.32128	0.32392	0.32656	0.32921
0.63	0.33185	0.33450	0.33716	0.33981	0.34247	0.34513	0.34779	0.35045	0.35312	0.35579
0.64	0.35846	0.36113	0.36381	0.36649	0.36917	0.37186	0.37454	0.37723	0.37993	0.38262
0.65	0.38532	0.38802	0.39073	0.39343	0.39614	0.39886	0.40157	0.40429	0.40701	0.40974
0.66	0.41246	0.41519	0.41793	0.42066	0.42341	0.42615	0.42889	0.43164	0.43440	0.43715
0.67	0.43991	0.44268	0.44544	0.44821	0.45099	0.45376	0.45654	0.45933	0.46211	0.46490
0.68	0.46770	0.47050	0.47330	0.47610	0.47891	0.48173	0.48454	0.48736	0.49019	0.49302
0.69	0.49585	0.49869	0.50153	0.50437	0.50722	0.51007	0.51293	0.51579	0.51866	0.52153
0.70	0.52440	0.52728	0.53016	0.53305	0.53594	0.53884	0.54174	0.54464	0.54755	0.55046
0.71	0.55338	0.55631	0.55924	0.56217	0.56511	0.56805	0.57100	0.57395	0.57691	0.57987
0.72	0.58284	0.58581	0.58879	0.59178	0.59477	0.59776	0.60076	0.60376	0.60678	0.60979
0.73	0.61281	0.61584	0.61887	0.62191	0.62496	0.62801	0.63106	0.63412	0.63719	0.64027
0.74	0.64334	0.64643	0.64952	0.65262	0.65573	0.65884	0.66196	0.66508	0.66821	0.67135
0.75	0.67449	0.67764	0.68080	0.68396	0.68713	0.69031	0.69349	0.69668	0.69988	0.70309
0.76	0.70630	0.70952	0.71275	0.71599	0.71923	0.72248	0.72574	0.72900	0.73228	0.73556
0.77	0.73885	0.74214	0.74545	0.74876	0.75208	0.75541	0.75875	0.76210	0.76546	0.76882
0.78	0.77219	0.77557	0.77897	0.78237	0.78577	0.78919	0.79262	0.79606	0.79950	0.80296
0.79	0.80642	0.80990	0.81338	0.81687	0.82038	0.82389	0.82742	0.83095	0.83450	0.83805
0.80	0.84162	0.84520	0.84879	0.85239	0.85600	0.85962	0.86325	0.86689	0.87055	0.87422
0.81	0.87790	0.88159	0.88529	0.88901	0.89273	0.89647	0.90023	0.90399	0.90777	0.91156
0.82	0.91537	0.91918	0.92301	0.92686	0.93072	0.93459	0.93848	0.94238	0.94629	0.95022
0.83	0.95416	0.95813	0.96210	0.96609	0.97009	0.97411	0.97815	0.98220	0.98627	0.99036
0.84	0.99446	0.99858	1.00271	1.00687	1.01104	1.01522	1.01943	1.02365	1.02789	1.03215
0.85	1.03643	1.04073	1.04505	1.04939	1.05375	1.05812	1.06252	1.06694	1.07138	1.07584
0.86	1.08032	1.08482	1.08935	1.09390	1.09847	1.10306	1.10768	1.11232	1.11699	1.12168
0.87	1.12639	1.13113	1.13590	1.14069	1.14550	1.15035	1.15522	1.16012	1.16505	1.17000
0.88	1.17499	1.18000	1.18504	1.19012	1.19522	1.20036	1.20553	1.21073	1.21596	1.22123
0.89	1.22653	1.23187	1.23724	1.24264	1.24809	1.25357	1.25908	1.26464	1.27024	1.27588
0.90	1.28155	1.28727	1.29303	1.29884	1.30469	1.31058	1.31652	1.32251	1.32854	1.33462
0.91	1.34075	1.34694	1.35317	1.35946	1.36581	1.37220	1.37866	1.38517	1.39175	1.39838
0.92	1.40507	1.41183	1.41865	1.42554	1.43250	1.43953	1.44663	1.45380	1.46106	1.46838
0.93	1.47579	1.48328	1.49085	1.49852	1.50626	1.51410	1.52203	1.53007	1.53820	1.54643
0.94	1.55477	1.56322	1.57179	1.58047	1.58927	1.59819	1.60725	1.61644	1.62576	1.63524
0.95	1.64485	1.65463	1.66456	1.67466	1.68494	1.69540	1.70604	1.71688	1.72793	1.73920
0.96	1.75069	1.76241	1.77438	1.78661	1.79912	1.81191	1.82501	1.83843	1.85218	1.86629
0.97	1.88079	1.89570	1.91103	1.92684	1.94314	1.95996	1.97737	1.99539	2.01409	2.03352
0.98	2.05375	2.07485	2.09693	2.12007	2.14441	2.17009	2.19728	2.22621	2.25713	2.29036
0.99	2.32634	2.36561	2.40892	2.45727	2.51213	2.57583	2.65209	2.74777	2.87815	3.09024

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg

<alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004



Fractiles de la t-Student ($t_{\alpha;v}$)

Dada $X:T(v)$, $t_{\alpha;v} = x$ tal que $P(X \leq x) = \alpha$

- Si $\alpha < 0.5$, se puede usar $t_{\alpha;v} = -t_{1-\alpha;v}$
- Si $v > 50$, se puede usar $t_{\alpha;v} \approx z_{\alpha}$

Ejemplos:

- $t_{0.75;10} = 0.6998$
- $t_{0.02;15} = -t_{0.98;15} = -2.2485$
- $t_{0.95;70} \approx z_{0.95} = 1.64485$

v	α						
	0.75	0.9	0.95	0.975	0.98	0.99	0.995
1	1.0000	3.0777	6.3137	12.7062	15.8945	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	4.8487	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	3.4819	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	2.9985	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	2.7565	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	2.6122	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.5168	2.9979	3.4995

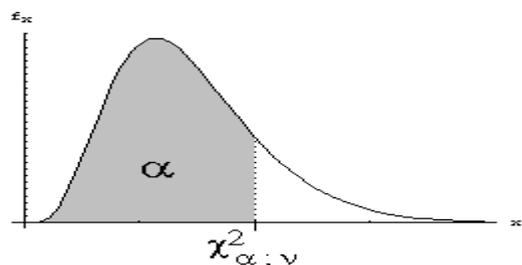
8	0.7064	1.3968	1.8595	2.3060	2.4490	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.3984	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.3593	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.3281	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.3027	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.2816	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.2638	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.2485	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.2354	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.2238	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.2137	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.2047	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.1967	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.1894	2.5176	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.1829	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.1770	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.1715	2.4922	2.7970
25	0.6844	1.3163	1.7081	2.0595	2.1666	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.1620	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.1578	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.1539	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.1503	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.1470	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.1438	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.1409	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.1382	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.1356	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.1332	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.1309	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.1287	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.1267	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.1247	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.1229	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.1212	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.1195	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.1179	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.1164	2.4141	2.6923
45	0.6800	1.3007	1.6794	2.0141	2.1150	2.4121	2.6896
46	0.6799	1.3002	1.6787	2.0129	2.1136	2.4102	2.6870
47	0.6797	1.2998	1.6779	2.0117	2.1123	2.4083	2.6846
48	0.6796	1.2994	1.6772	2.0106	2.1111	2.4066	2.6822
49	0.6795	1.2991	1.6766	2.0096	2.1099	2.4049	2.6800
50	0.6794	1.2987	1.6759	2.0086	2.1087	2.4033	2.6778

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg

<alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004



Fractiles de la chi-cuadrada ($\chi_{\alpha;v}$)

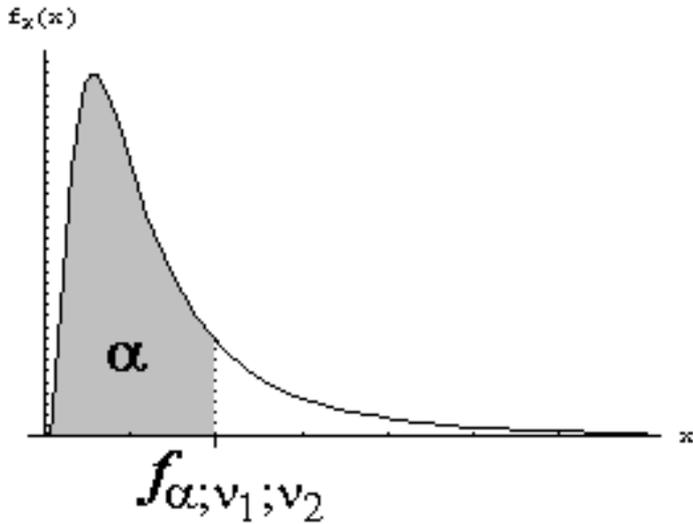
Dada $X: \chi^2(v)$, $\chi^2_{\alpha;v} = x$ tal que $P(X \leq x) = \alpha$

Ejemplo:

- $\chi^2_{0.75;10} = 12.549$

	0.005	0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99	0.995
1	3.927 10 ⁻⁵	1.571 10 ⁻⁴	9.821 10 ⁻⁴	3.932 10 ⁻³	0.0158	0.102	0.455	1.323	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	7.584	10.341	13.701	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	8.438	11.340	14.845	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	9.299	12.340	15.984	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	10.165	13.339	17.117	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	11.037	14.339	18.245	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	11.912	15.338	19.369	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	12.792	16.338	20.489	24.769	27.587	30.191	33.409	35.718

18	6.265	7.015	8.231	9.390	10.865	13.675	17.338	21.605	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	14.562	18.338	22.718	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	15.452	19.337	23.828	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	16.344	20.337	24.935	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	17.240	21.337	26.039	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	18.137	22.337	27.141	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	19.037	23.337	28.241	33.196	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	16.473	19.939	24.337	29.339	34.382	37.652	40.646	44.314	46.928
30	13.787	14.953	16.791	18.493	20.599	24.478	29.336	34.800	40.256	43.773	46.979	50.892	53.672
35	17.192	18.509	20.569	22.465	24.797	29.054	34.336	40.223	46.059	49.802	53.203	57.342	60.275
40	20.707	22.164	24.433	26.509	29.051	33.660	39.335	45.616	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	42.942	49.335	56.334	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	52.294	59.335	66.981	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	61.698	69.334	77.577	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	71.145	79.334	88.130	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	80.625	89.334	98.650	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	90.133	99.334	109.141	118.498	124.342	129.561	135.807	140.170
200	152.24	156.43	162.73	168.28	174.84	186.17	199.33	213.10	226.02	233.99	241.06	249.45	255.26
300	240.66	245.97	253.91	260.88	269.07	283.14	299.33	316.14	331.79	341.40	349.87	359.91	366.84
400	330.90	337.16	346.48	354.64	364.21	380.58	399.33	418.70	436.65	447.63	457.31	468.72	476.61
500	422.30	429.39	439.94	449.15	459.93	478.32	499.33	520.95	540.93	553.13	563.85	576.49	585.21
600	514.53	522.37	534.02	544.18	556.06	576.29	599.33	622.99	644.80	658.09	669.77	683.52	692.98
700	607	616	629	640	652	674	699	725	748	763	775	790	800
800	700	710	724	735	749	773	799	827	852	867	880	896	907
900	794	804	819	831	846	871	899	928	955	971	985	1002	1013
1000	889	899	914	928	943	969	999	1030	1058	1075	1090	1107	1119
2000	1841	1856	1878	1897	1919	1957	1999	2042	2081	2105	2126	2150	2167
5000	4746	4770	4806	4837	4872	4932	4999	5067	5129	5166	5198	5236	5261
10000	9639	9674	9724	9769	9819	9904	9999	10095	10182	10234	10279	10331	10368



Fractiles de la F ($f_{\alpha; v_1; v_2}$)

Dada X: $F(v_1, v_2)$, $f_{\alpha; v_1; v_2} = x$ tal que $P(X \leq x) = \alpha$

A continuación damos 4 tablas, para $\alpha = 0.9, 0.95, 0.99, 0.995$. Esos son valores de α cercanos a 1.

$$f_{\alpha; v_1; v_2} = \frac{1}{f_{1-\alpha; v_2; v_1}}$$

En caso de necesitarse los fractiles para α cercano a cero, se puede usar:

Ejemplos:

- $f_{0.9; 4; 8} = 2.806$
- $f_{0.1; 3; 7} = 1 / f_{0.9; 7; 3} = 1 / 5.266 = 0.19$

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final. Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 1 de junio de 2004_

Fractiles de la F ($f_{\alpha; v_1; v_2}$)

Dada X: F(v_1, v_2), $f_{\alpha; v_1; v_2} = x$ tal que $P(X \leq x) = \alpha$

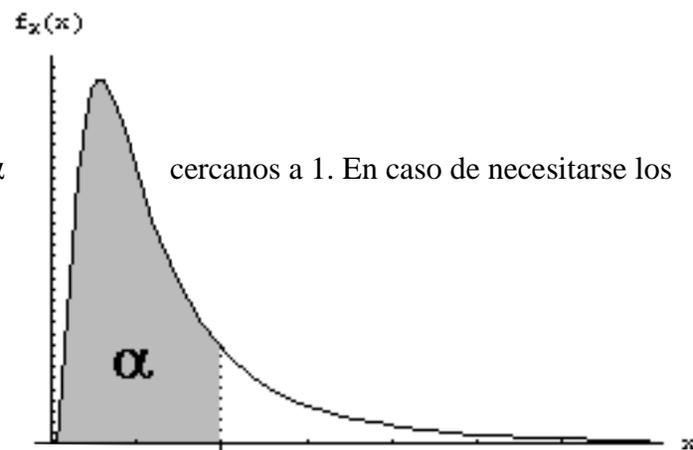
A continuación damos 4 tablas, para $\alpha = 0.9, 0.95, 0.99, 0.995$. Esos son valores de α cercanos a 1. En caso de necesitarse los fractiles para α cercano a cero, se puede usar:

$f_{\alpha; v_1; v_2} = \frac{1}{f_{1-\alpha; v_1; v_2}}$. Ejemplos:

- $f_{0.9; 4; 8} = 2.806$
- $f_{0.1; 3; 7} = 1 / f_{0.9; 7; 3} = 1 / 5.266 = 0.19$

Tabla para $\alpha = 0.9$:

v_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$f_{\alpha; v_1; v_2}$					40	50	100	500	1000
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.47	60.71	60.90	61.07	61.22	61.74	62.05	62.26	62.53	62.69	63.01	63.26	63.30		
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392	9.401	9.408	9.415	9.420	9.425	9.441	9.451	9.458	9.466	9.471	9.481	9.489	9.490		
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230	5.222	5.216	5.210	5.205	5.200	5.184	5.175	5.168	5.160	5.155	5.144	5.136	5.135		
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920	3.907	3.896	3.886	3.878	3.870	3.844	3.828	3.817	3.804	3.795	3.778	3.764	3.762		
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297	3.282	3.268	3.257	3.247	3.238	3.207	3.187	3.174	3.157	3.147	3.126	3.109	3.107		
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937	2.920	2.905	2.892	2.881	2.871	2.836	2.815	2.800	2.781	2.770	2.746	2.727	2.725		
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703	2.684	2.668	2.654	2.643	2.632	2.595	2.571	2.555	2.535	2.523	2.497	2.476	2.473		
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538	2.519	2.502	2.488	2.475	2.464	2.425	2.400	2.383	2.361	2.348	2.321	2.298	2.295		
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416	2.396	2.379	2.364	2.351	2.340	2.298	2.272	2.255	2.232	2.218	2.189	2.165	2.162		
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323	2.302	2.284	2.269	2.255	2.244	2.201	2.174	2.155	2.132	2.117	2.087	2.062	2.059		
11	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274	2.248	2.227	2.209	2.193	2.179	2.167	2.123	2.095	2.076	2.052	2.036	2.005	1.979	1.975		
12	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214	2.188	2.166	2.147	2.131	2.117	2.105	2.060	2.031	2.011	1.986	1.970	1.938	1.911	1.907		
13	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164	2.138	2.116	2.097	2.080	2.066	2.053	2.007	1.978	1.958	1.931	1.915	1.882	1.853	1.850		
14	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122	2.095	2.073	2.054	2.037	2.022	2.010	1.962	1.933	1.912	1.885	1.869	1.834	1.805	1.801		
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059	2.037	2.017	2.000	1.985	1.972	1.924	1.894	1.873	1.845	1.828	1.793	1.763	1.759		
16	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055	2.028	2.005	1.985	1.968	1.953	1.940	1.891	1.860	1.839	1.811	1.793	1.757	1.726	1.722		
17	3.026	2.645	2.437	2.308	2.218	2.152	2.102	2.061	2.028	2.001	1.978	1.958	1.940	1.925	1.912	1.862	1.831	1.809	1.781	1.763	1.726	1.694	1.690		
18	3.007	2.624	2.416	2.286	2.196	2.130	2.079	2.038	2.005	1.977	1.954	1.933	1.916	1.900	1.887	1.837	1.805	1.783	1.754	1.736	1.698	1.665	1.661		
19	2.990	2.606	2.397	2.266	2.176	2.109	2.058	2.017	1.984	1.956	1.932	1.912	1.894	1.878	1.865	1.814	1.782	1.759	1.730	1.711	1.673	1.639	1.635		
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937	1.913	1.892	1.875	1.859	1.845	1.794	1.761	1.738	1.708	1.690	1.650	1.616	1.612		
25	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866	1.841	1.820	1.802	1.785	1.771	1.718	1.683	1.659	1.627	1.607	1.565	1.527	1.523		
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819	1.794	1.773	1.754	1.737	1.722	1.667	1.632	1.606	1.573	1.552	1.507	1.467	1.462		
35	2.855	2.461	2.247	2.113	2.019	1.950	1.896	1.852	1.817	1.787	1.761	1.739	1.720	1.703	1.688	1.632	1.595	1.569	1.535	1.513	1.465	1.423	1.417		
40	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793	1.763	1.737	1.715	1.695	1.678	1.662	1.605	1.568	1.541	1.506	1.483	1.434	1.389	1.383		



70	2.779	2.380	2.164	2.027	1.931	1.860	1.804	1.760	1.723	1.691	1.665	1.641	1.621	1.603	1.587	1.526	1.486	1.457	1.418	1.392	1.335	1.281	1.273
100	2.756	2.356	2.139	2.002	1.906	1.834	1.778	1.732	1.695	1.663	1.636	1.612	1.592	1.573	1.557	1.494	1.453	1.423	1.382	1.355	1.293	1.232	1.223
200	2.731	2.329	2.111	1.973	1.876	1.804	1.747	1.701	1.663	1.631	1.603	1.579	1.558	1.539	1.522	1.458	1.414	1.383	1.339	1.310	1.242	1.168	1.157
500	2.716	2.313	2.095	1.956	1.859	1.786	1.729	1.683	1.644	1.612	1.583	1.559	1.537	1.518	1.501	1.435	1.391	1.358	1.313	1.282	1.209	1.122	1.106
1000	2.711	2.308	2.089	1.950	1.853	1.780	1.723	1.676	1.638	1.605	1.577	1.552	1.531	1.511	1.494	1.428	1.383	1.350	1.304	1.273	1.197	1.103	1.084

Tabla para $\alpha = 0.95$:

v_2	v_1																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	25	30	40	50	100	500	1000
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.90	244.69	245.36	245.95	248.02	249.26	250.10	251.14	251.77	253.04	254.06	254.19
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396	19.405	19.412	19.419	19.424	19.429	19.446	19.456	19.463	19.471	19.476	19.486	19.494	19.495
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.763	8.745	8.729	8.715	8.703	8.660	8.634	8.617	8.594	8.581	8.554	8.532	8.529
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.936	5.912	5.891	5.873	5.858	5.803	5.769	5.746	5.717	5.699	5.664	5.635	5.632
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.704	4.678	4.655	4.636	4.619	4.558	4.521	4.496	4.464	4.444	4.405	4.373	4.369
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.976	3.956	3.938	3.874	3.835	3.808	3.774	3.754	3.712	3.678	3.673
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.550	3.529	3.511	3.445	3.404	3.376	3.340	3.319	3.275	3.239	3.234
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.259	3.237	3.218	3.150	3.108	3.079	3.043	3.020	2.975	2.937	2.932
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.048	3.025	3.006	2.936	2.893	2.864	2.826	2.803	2.756	2.717	2.712
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.887	2.865	2.845	2.774	2.730	2.700	2.661	2.637	2.588	2.548	2.543
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.761	2.739	2.719	2.646	2.601	2.570	2.531	2.507	2.457	2.415	2.410
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.660	2.637	2.617	2.544	2.498	2.466	2.426	2.401	2.350	2.307	2.302
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.577	2.554	2.533	2.459	2.412	2.380	2.339	2.314	2.261	2.218	2.212
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.507	2.484	2.463	2.388	2.341	2.308	2.266	2.241	2.187	2.142	2.136
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.448	2.424	2.403	2.328	2.280	2.247	2.204	2.178	2.123	2.078	2.072
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.456	2.425	2.397	2.373	2.352	2.276	2.227	2.194	2.151	2.124	2.068	2.022	2.016
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.413	2.381	2.353	2.329	2.308	2.230	2.181	2.148	2.104	2.077	2.020	1.973	1.967
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.374	2.342	2.314	2.290	2.269	2.191	2.141	2.107	2.063	2.035	1.978	1.929	1.923
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.340	2.308	2.280	2.256	2.234	2.155	2.106	2.071	2.026	1.999	1.940	1.891	1.884
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.310	2.278	2.250	2.225	2.203	2.124	2.074	2.039	1.994	1.966	1.907	1.856	1.850
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.198	2.165	2.136	2.111	2.089	2.007	1.955	1.919	1.872	1.842	1.779	1.725	1.718
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.126	2.092	2.063	2.037	2.015	1.932	1.878	1.841	1.792	1.761	1.695	1.637	1.630
35	4.121	3.267	2.874	2.641	2.485	2.372	2.285	2.217	2.161	2.114	2.075	2.041	2.012	1.986	1.963	1.878	1.824	1.786	1.735	1.703	1.635	1.574	1.566
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.038	2.003	1.974	1.948	1.924	1.839	1.783	1.744	1.693	1.660	1.589	1.526	1.517
70	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969	1.928	1.893	1.863	1.836	1.812	1.722	1.664	1.622	1.566	1.530	1.450	1.374	1.364
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.886	1.850	1.819	1.792	1.768	1.676	1.616	1.573	1.515	1.477	1.392	1.308	1.296
200	3.888	3.041	2.650	2.417	2.259	2.144	2.056	1.985	1.927	1.878	1.837	1.801	1.769	1.742	1.717	1.623	1.561	1.516	1.455	1.415	1.321	1.221	1.205
500	3.860	3.014	2.623	2.390	2.232	2.117	2.028	1.957	1.899	1.850	1.808	1.772	1.740	1.712	1.686	1.592	1.528	1.482	1.419	1.376	1.275	1.159	1.138
1000	3.851	3.005	2.614	2.381	2.223	2.108	2.019	1.948	1.889	1.840	1.798	1.762	1.730	1.702	1.676	1.581	1.517	1.471	1.406	1.363	1.260	1.134	1.110

Tabla para $\alpha = 0.99$:

v_2	v_1																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	25	30	40	50	100	500	1000
1	4052.1	4999.3	5403.5	5624.2	5763.9	5858.9	5928.3	5980.9	6022.4	6055.9	6083.4	6106.6	6125.7	6143.0	6156.9	6208.6	6239.8	6260.3	6286.4	6302.2	6333.9	6359.5	6362.8

2	98.502	99.000	99.164	99.251	99.302	99.331	99.357	99.375	99.390	99.397	99.408	99.419	99.422	99.426	99.433	99.448	99.459	99.466	99.477	99.477	99.491	99.499	99.499
3	34.116	30.816	29.457	28.710	28.237	27.911	27.671	27.489	27.345	27.228	27.132	27.052	26.983	26.924	26.872	26.690	26.579	26.504	26.411	26.354	26.241	26.148	26.137
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.452	14.374	14.306	14.249	14.198	14.019	13.911	13.838	13.745	13.690	13.577	13.486	13.475
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.963	9.888	9.825	9.770	9.722	9.553	9.449	9.379	9.291	9.238	9.130	9.042	9.032
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.790	7.718	7.657	7.605	7.559	7.396	7.296	7.229	7.143	7.091	6.987	6.901	6.891
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.538	6.469	6.410	6.359	6.314	6.155	6.058	5.992	5.908	5.858	5.755	5.671	5.660
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.734	5.667	5.609	5.559	5.515	5.359	5.263	5.198	5.116	5.065	4.963	4.880	4.869
9	10.562	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.178	5.111	5.055	5.005	4.962	4.808	4.713	4.649	4.567	4.517	4.415	4.332	4.321
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.772	4.706	4.650	4.601	4.558	4.405	4.311	4.247	4.165	4.115	4.014	3.930	3.920
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.462	4.397	4.342	4.293	4.251	4.099	4.005	3.941	3.860	3.810	3.708	3.624	3.613
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.220	4.155	4.100	4.052	4.010	3.858	3.765	3.701	3.619	3.569	3.467	3.382	3.372
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	4.025	3.960	3.905	3.857	3.815	3.665	3.571	3.507	3.425	3.375	3.272	3.187	3.176
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.864	3.800	3.745	3.698	3.656	3.505	3.412	3.348	3.266	3.215	3.112	3.026	3.015
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.730	3.666	3.612	3.564	3.522	3.372	3.278	3.214	3.132	3.081	2.977	2.891	2.880
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.616	3.553	3.498	3.451	3.409	3.259	3.165	3.101	3.018	2.967	2.863	2.775	2.764
17	8.400	6.112	5.185	4.669	4.336	4.101	3.927	3.791	3.682	3.593	3.518	3.455	3.401	3.353	3.312	3.162	3.068	3.003	2.920	2.869	2.764	2.676	2.664
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.434	3.371	3.316	3.269	3.227	3.077	2.983	2.919	2.835	2.784	2.678	2.589	2.577
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.360	3.297	3.242	3.195	3.153	3.003	2.909	2.844	2.761	2.709	2.602	2.512	2.501
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.294	3.231	3.177	3.130	3.088	2.938	2.843	2.778	2.695	2.643	2.535	2.445	2.433
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	3.056	2.993	2.939	2.892	2.850	2.699	2.604	2.538	2.453	2.400	2.289	2.194	2.182
30	7.562	5.390	4.510	4.018	3.699	3.473	3.305	3.173	3.067	2.979	2.906	2.843	2.789	2.742	2.700	2.549	2.453	2.386	2.299	2.245	2.131	2.032	2.019
35	7.419	5.268	4.396	3.908	3.592	3.368	3.200	3.069	2.963	2.876	2.803	2.740	2.686	2.639	2.597	2.445	2.348	2.281	2.193	2.137	2.020	1.918	1.905
40	7.314	5.178	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.727	2.665	2.611	2.563	2.522	2.369	2.271	2.203	2.114	2.058	1.938	1.833	1.819
70	7.011	4.922	4.074	3.600	3.291	3.071	2.906	2.777	2.672	2.585	2.512	2.450	2.395	2.348	2.306	2.150	2.050	1.980	1.886	1.826	1.695	1.574	1.558
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503	2.430	2.368	2.313	2.265	2.223	2.067	1.965	1.893	1.797	1.735	1.598	1.466	1.447
200	6.763	4.713	3.881	3.414	3.110	2.893	2.730	2.601	2.497	2.411	2.338	2.275	2.220	2.172	2.129	1.971	1.868	1.794	1.694	1.629	1.481	1.328	1.304
500	6.686	4.648	3.821	3.357	3.054	2.838	2.675	2.547	2.443	2.356	2.283	2.220	2.166	2.117	2.075	1.915	1.810	1.735	1.633	1.566	1.408	1.232	1.201
1000	6.660	4.626	3.801	3.338	3.036	2.820	2.657	2.529	2.425	2.339	2.265	2.203	2.148	2.099	2.056	1.897	1.791	1.716	1.613	1.544	1.383	1.195	1.159

Tabla para $\alpha = 0.995$:

v_2	v_1																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	25	30	40	50	100	500	1000
1	16212	19997	21614	22500	23055	23439	23715	23923	24091	24221	24333	24426	24505	24572	24631	24836	24959	25041	25145	25212	25339	25436	25451
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5
3	55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.88	43.68	43.52	43.39	43.27	43.17	43.08	42.78	42.59	42.47	42.31	42.21	42.02	41.87	41.85
4	31.33	26.28	24.26	23.15	22.46	21.98	21.62	21.35	21.14	20.97	20.82	20.70	20.60	20.51	20.44	20.17	20.00	19.89	19.75	19.67	19.50	19.36	19.34
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.49	13.38	13.29	13.21	13.15	12.90	12.76	12.66	12.53	12.45	12.30	12.17	12.16
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.13	10.03	9.95	9.88	9.81	9.59	9.45	9.36	9.24	9.17	9.03	8.91	8.89
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.27	8.18	8.10	8.03	7.97	7.75	7.62	7.53	7.42	7.35	7.22	7.10	7.09
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.10	7.01	6.94	6.87	6.81	6.61	6.48	6.40	6.29	6.22	6.09	5.98	5.96
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.31	6.23	6.15	6.09	6.03	5.83	5.71	5.62	5.52	5.45	5.32	5.21	5.20
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.75	5.66	5.59	5.53	5.47	5.27	5.15	5.07	4.97	4.90	4.77	4.67	4.65
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.32	5.24	5.16	5.10	5.05	4.86	4.74	4.65	4.55	4.49	4.36	4.25	4.24
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.99	4.91	4.84	4.77	4.72	4.53	4.41	4.33	4.23	4.17	4.04	3.93	3.92
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.72	4.64	4.57	4.51	4.46	4.27	4.15	4.07	3.97	3.91	3.78	3.67	3.66

14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.51	4.43	4.36	4.30	4.25	4.06	3.94	3.86	3.76	3.70	3.57	3.46	3.45
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.33	4.25	4.18	4.12	4.07	3.88	3.77	3.69	3.59	3.52	3.39	3.29	3.27
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.18	4.10	4.03	3.97	3.92	3.73	3.62	3.54	3.44	3.37	3.25	3.14	3.13
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	4.05	3.97	3.90	3.84	3.79	3.61	3.49	3.41	3.31	3.25	3.12	3.01	3.00
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.94	3.86	3.79	3.73	3.68	3.50	3.38	3.30	3.20	3.14	3.01	2.90	2.89
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.84	3.76	3.70	3.64	3.59	3.40	3.29	3.21	3.11	3.04	2.91	2.80	2.79
20	9.944	6.987	5.818	5.174	4.762	4.472	4.257	4.090	3.956	3.847	3.756	3.678	3.611	3.553	3.502	3.318	3.203	3.123	3.022	2.959	2.828	2.719	2.705
25	9.475	6.598	5.462	4.835	4.433	4.150	3.939	3.776	3.645	3.537	3.447	3.370	3.304	3.247	3.196	3.013	2.898	2.819	2.716	2.652	2.519	2.406	2.391
30	9.180	6.355	5.239	4.623	4.228	3.949	3.742	3.580	3.451	3.344	3.255	3.179	3.113	3.056	3.006	2.823	2.708	2.628	2.524	2.459	2.323	2.207	2.191
35	8.976	6.188	5.086	4.479	4.088	3.812	3.607	3.447	3.318	3.212	3.124	3.048	2.983	2.926	2.876	2.693	2.577	2.497	2.392	2.327	2.188	2.068	2.052
40	8.828	6.066	4.976	4.374	3.986	3.713	3.509	3.350	3.222	3.117	3.028	2.953	2.888	2.831	2.781	2.598	2.482	2.401	2.296	2.230	2.088	1.965	1.948
70	8.403	5.720	4.661	4.076	3.698	3.431	3.232	3.076	2.950	2.846	2.759	2.684	2.619	2.563	2.513	2.329	2.211	2.128	2.019	1.949	1.797	1.657	1.637
100	8.241	5.589	4.542	3.963	3.589	3.325	3.127	2.972	2.847	2.744	2.657	2.583	2.518	2.461	2.411	2.227	2.108	2.024	1.912	1.840	1.681	1.529	1.508
200	8.057	5.441	4.408	3.837	3.467	3.206	3.010	2.856	2.732	2.629	2.543	2.468	2.404	2.347	2.297	2.112	1.991	1.905	1.790	1.715	1.544	1.369	1.343
500	7.950	5.355	4.330	3.763	3.396	3.137	2.941	2.789	2.665	2.562	2.476	2.402	2.337	2.281	2.230	2.044	1.922	1.835	1.717	1.640	1.460	1.260	1.225
1000	7.914	5.326	4.305	3.739	3.373	3.114	2.919	2.766	2.643	2.541	2.454	2.380	2.315	2.259	2.208	2.022	1.900	1.812	1.693	1.615	1.431	1.218	1.177

El siguiente material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 11 de junio de 2004

APÉNDICE E

Resumen de fórmulas de probabilidad

A continuación se presenta un resumen de las fórmulas que aparecen en los capítulos dedicados a probabilidad (capítulos 1 a 7). El mismo no incluye las fórmulas usadas en estadística (capítulos 8 a 11) ni las que aparecen en los demás apéndices.

Fórmulas básicas de probabilidad (Capítulo 1)

Definición de Laplace (Sección 1.2)

$$P(A) = \frac{\text{cantidad de resultados contenidos en } A}{\text{cantidad total de resultados}}$$

Definición empírica (Sección 1.2)

$$P(A) \approx \text{fr}_{\text{rel}}(A) = \frac{\text{fr}_{\text{abs}}(A)}{n}$$

Axiomas y consecuencias (Sección 1.2)

- $P(A) \geq 0$
- $P(E) = 1$
- $A \cap B = \emptyset \Leftrightarrow P(A \cup B) = P(A) + P(B)$
- $P(A) \leq 1$
- $P(A) + P(\bar{A}) = 1$
- $P(\emptyset) = 0$
- $A \subset B \Rightarrow P(A) \leq P(B)$

Suma de probabilidades (Sección 1.2)

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

Probabilidad condicional (Sección 1.3)

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

Multiplicación de probabilidades (Sección 1.3)

$$P(A \cap B \cap C) = P(A) P(B/A) P(C/A \cap B)$$
$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P\left(A_i / \bigcap_{j=1}^{i-1} A_j\right)$$

Independencia de sucesos (Sección 1.4)

- A, B indep. $\Leftrightarrow P(A/B) = P(A) \Leftrightarrow P(B/A) = P(B) \Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$
- A, B indep. $\Leftrightarrow A, B^c$ indep. $\Leftrightarrow A^c, B$ indep. $\Leftrightarrow A^c, B^c$ indep.

Probabilidad total (Sección 1.5)

$$P(A) = \sum_{i=1}^n P(A \cap p_i) = \sum_{i=1}^n P(A/p_i) \cdot P(p_i)$$

Regla de Bayes (Sección 1.6)

$$P(p_i/A) = \frac{P(A/p_i) P(p_i)}{\sum_{i=1}^n P(A/p_i) P(p_i)}$$

Variables aleatorias unidimensionales (Capítulo 2)

Funciones de densidad y distribución y probabilidades (Sección 2.3)

$$P(X \leq x_0) = F_X(x_0) = \sum_{x=-\infty}^{x_0} P_X(x)$$

(X discreta)

$$P(X \leq x_0) = P(X < x_0) = F_X(x_0) = \int_{-\infty}^{x_0} f_X(x) dx$$

(X continua)

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

Cambio de variables continuo (Sección 2.4)

$$f_Y(y) = \frac{f_X(x)}{\left| \frac{dy}{dx} \right|}$$

Esperanza (Sección 2.5)

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

•

$$E(\varphi(x)) = \int_{-\infty}^{+\infty} \varphi(x) f_X(x) dx$$

•

• Para X discreta, reemplazar integrales por sumatorias y f_X por P_X .

$$E(aX + b) = E(aX) + E(b) = aE(X) + b \quad \text{con } a, b \in \mathfrak{R}$$

Varianza (Sección 2.6)

$$Var(X) = \sigma_X^2 = E((X - E(X))^2) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx$$

•

$$\sigma_X^2 = E(X^2) - E(X)^2$$

•

$$\sigma^2(aX + b) = a^2 \sigma_X^2 \quad \text{con } a, b \in \mathfrak{R}$$

•

Mezcla (Sección 2.9)

$$f_{X_{MEZCLA}} = P(A_1) f_{X_1}(x) + P(A_2) f_{X_2}(x) + \dots + P(A_n) f_{X_n}(x)$$

Variables aleatorias bidimensionales y n-dimensionales (Capítulo 3)

Marginación (Sección 3.3)

$$P_X(x) = \sum_{y=-\infty}^{+\infty} P_{XY}(x, y)$$

•

para variables discretas

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

•

para variables continuas

Distribución condicional (Sección 3.4)

$$P_{X/Y}(x, y) = \frac{P_{XY}(x, y)}{P_Y(y)}$$

•

para variables discretas

$$f_{X/Y}(x, y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

•

para variables continuas

Independencia de variables aleatorias (Sección 3.5)

$$X \text{ e } Y \text{ indep.} \Leftrightarrow f_{XY}(x, y) = f_X(x) \Leftrightarrow f_{YX}(x, y) = f_Y(y) \Leftrightarrow f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$$

- Para variables discretas es análogo

Esperanza condicional (Sección 3.6)

$$E(X/Y) = \mu_{X/Y} = \int_{-\infty}^{\infty} x f_{X/Y}(x, y) dx$$

-
- Para variables discretas es análogo

Cambio de variables (Sección 3.7 , 3.8)

$$f_{UV}(u, v) = \frac{f_{XY}(x, y)}{\left| \frac{\partial(u, v)}{\partial(x, y)} \right|}$$

-
- $E(\varphi(x, y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x, y) f_{XY}(x, y) dy dx$

- $E(X + Y) = E(X) + E(Y)$

- $E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i)$

- $\sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$

- $\text{cov}(X, Y) = \sigma_{XY} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y) f_{XY}(x, y) dy dx = E(XY) - \mu_X \mu_Y$

- $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

Máximos y mínimos (Sección 3.9)

Hipótesis sobre las variables aleatorias X_i :	$Y = \max\{X_1, X_2, \dots, X_n\}$	$Y = \min\{X_1, X_2, \dots, X_n\}$
Las X_i son independientes e idénticamente distribuidas	$f_Y(y) = n [F_X(y)]^{n-1} f_X(y)$ $F_Y(y) = [F_X(y)]^n$	$f_Y(y) = n [1 - F_X(y)]^{n-1} f_X(y)$ $F_Y(y) = 1 - [1 - F_X(y)]^n$
Las X_i son independientes, y cada una tiene su propia distribución	$F_Y(y) = \prod_{i=1}^n [F_{X_i}(y)] =$ $= [F_{X_1}(y)] \dots [F_{X_n}(y)]$	$F_Y(y) = 1 - \prod_{i=1}^n [1 - F_{X_i}(y)] =$ $= 1 - [1 - F_{X_1}(y)] \dots [1 - F_{X_n}(y)]$

Las X_i no son independientes

$$F_Y(y) =$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1 X_2 \dots X_n} dx_n \dots dx_2 dx_1$$

$$F_Y(y) = 1 -$$

$$\int_y^{\infty} \int_y^{\infty} \dots \int_y^{\infty} f_{X_1 X_2 \dots X_n} dx_n \dots dx_2 dx_1$$

Distribuciones particulares (Capítulos 4 - 7)

Nombre	Cap.	Función de probabilidad / densidad	Esperanza	Varianza
Beta	7	$f_x(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} & 0 < x < 1 \\ 0 & \forall \text{ otro } x \end{cases}$ <p>(***)</p>	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2 (a+b+1)}$
Binomial	4	$P_x(x) = \begin{cases} \binom{n}{x} p^x \cdot (1-p)^{n-x} & 0 \leq x \leq n \\ 0 & \forall \text{ otro } x \end{cases}$	n.p	n.p.(1-p)
Chi-cuadrada	7	(*)	v	2v
Exponencial negativa	5	$f_x(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$	1 / λ	1 / λ^2
F	7	(*)	(*)	(*)
Gamma (**)	5	$f_x(x) = \begin{cases} \frac{\lambda (\lambda x)^{k-1} e^{-\lambda x}}{\Gamma(k)} & x > 0 \\ 0 & x \leq 0 \end{cases}$ <p>(***)</p>	k / λ	k / λ^2
Geométrica	4	$P_x(x) = \begin{cases} p \cdot (1-p)^{x-1} & x \geq 1 \\ 0 & \forall \text{ otro } x \end{cases}$	1 / p	1 / p ²
Hipergeométrica	7	$P_x(x) = \frac{\binom{k}{x} \cdot \binom{N-k}{n-x}}{\binom{N}{n}}$	--	--
Multinomial	7	$P(X = x) = n! \prod_{i=1}^k \frac{p_i^{x_i}}{x_i!}$	--	--
Normal (ver aparte)	6	$f_x(x) = \frac{e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}}{\sqrt{2\pi} \sigma} \quad \forall x \in \mathfrak{R}$	μ	σ^2
Pascal	4	$P_x(x) = \begin{cases} \binom{x-1}{k-1} p^k \cdot (1-p)^{x-k} & x \geq k \\ 0 & \forall \text{ otro } x \end{cases}$	k / p	k / p ²

Poisson	5	$P_X(x) = \begin{cases} \frac{e^{-\mu} \mu^x}{x!} & x \geq 0 \\ 0 & x < 0 \end{cases}$	μ	μ
t-Student	7	(*)	0	$\frac{v}{v-2}$
Uniforme	7	$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \forall \text{ otro } x \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

(*) No resulta de utilidad

(**) Para calcular probabilidades de la gamma se puede usar:

$$\int_0^{x_0} f_X(x) dx = 1 - \sum_{i=0}^{x_0} P(Y=i)$$

$$\int_{x_0}^{+\infty} f_X(x) dx = \sum_{i=0}^{x_0} P(Y=i)$$

donde X:Gamma(λ, k) e Y:Poisson(μ) con $\mu = \lambda \cdot x_0$

$$\Gamma(k) = \int_0^{+\infty} x^{k-1} e^{-x} dx$$

(***)

Para k natural, vale $\Gamma(k) = (k-1)!$

Distribución normal (Sección 6.1)

$$Z = \frac{X - \mu}{\sigma} \Rightarrow Z: N(0,1)$$

$$P(X \leq x) = F_X(x) = F_Z\left(\frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

• Valores tabulados:

• Fractiles tabulados: Dada Z:N(0;1), $z_\alpha = z$ tal que $\Phi(z) = P(Z \leq z) = \alpha$

• Función lineal: X:N(μ_x ; σ_x) \wedge Y = aX+b \Rightarrow Y:N(a μ_x + b ; $\sigma_x |a|$)

$$Z = \sum_{i=1}^n \alpha_i X_i \Rightarrow$$

• Combinación lineal: $X_i:N(\mu_i; \sigma_i)$ independientes \wedge

$$Z : N\left(\mu_z = \sum_{i=1}^n \alpha_i \mu_i \quad ; \quad \sigma_z = \sqrt{\sum_{i=1}^n \alpha_i^2 \sigma_i^2}\right)$$

Teorema central del límite (Sección 6.2)

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

- tiene una distribución aproximadamente normal estándar

$$Y = \sum_{i=1}^n X_i$$

- tiene una distribución aproximadamente $N(n\mu ; \sqrt{n}\sigma)$

Este material se encuentra en etapa de corrección y no deberá ser considerado una versión final.

Para hacer comentarios y sugerencias, o reportar errores, enviar mail a Alejandro D. Zylberberg <alejandro@probabilidad.com.ar>

Versión Actualizada al: 6 de junio de 2004_