

Inferencia Estadística

Ms. Carlos López de Castilla Vásquez¹

20 de enero de 2012

¹Si tienes comentarios envíalos al siguiente correo electrónico:
`clopez@lamolina.edu.pe`

Índice general

1. Teoría de la probabilidad	1
1.1. Teoría de conjuntos	1
1.2. Fundamentos de la teoría de la probabilidad	4
1.2.1. Axiomas de la probabilidad	4
1.2.2. Cálculo de probabilidades	4
1.2.3. Conteo	5
1.2.4. Puntos igualmente probables	7
1.3. Probabilidad condicional e independencia	7
1.4. Variables aleatorias	8
1.5. Función de distribución acumulada	8
1.6. Función de probabilidad y densidad	9
2. Transformaciones y esperanza	10
2.1. Transformaciones para variables aleatorias	10
2.1.1. Caso discreto	11
2.1.2. Caso continuo	11
2.2. Valores esperados	16
2.3. Momentos y función generatriz de momentos	19
2.4. Derivación bajo el signo de la integral	26
3. Familias de distribuciones	31
3.1. Distribuciones discretas	31
3.2. Distribuciones continuas	36
3.3. Familias exponenciales	42
3.4. Familias de locación y escala	46
3.5. Desigualdades e identidades	48
3.5.1. Desigualdades para probabilidades	48
3.5.2. Identidades	48

4. Variables aleatorias múltiples	49
4.1. Distribución conjunta y marginal	49
4.2. Distribuciones condicionales e independencia	54
4.3. Transformaciones bivariadas	61
4.3.1. Caso discreto	61
4.3.2. Caso continuo	62
4.4. Modelos jerárquicos y distribuciones mixtas	66
4.5. Covarianza y correlación	71
4.6. Distribuciones multivariadas	75
4.7. Transformaciones sobre un vector aleatorio	80
4.8. Desigualdades	81
4.8.1. Desigualdades numéricas	81
5. Propiedades en una muestra aleatoria	83
5.1. Conceptos básicos sobre muestras aleatorias	83
5.2. Sumas de variables aleatorias a partir de una muestra aleatoria	84
5.3. Muestreo desde la distribución Normal	88
5.3.1. Propiedades de la media y variancia muestral	88
5.3.2. Distribuciones derivadas: t de Student y F de Snedecor	89
5.4. Estadísticas de orden	90
5.5. Conceptos de convergencia	94
5.5.1. Convergencia en probabilidad	94
5.5.2. Convergencia casi segura	95
5.5.3. Convergencia en distribución	96
6. Principios de reducción de la data	100
6.1. Introducción	100
6.2. El principio de suficiencia	100
6.2.1. Estadística suficiente	101
6.2.2. Estadística minimal suficiente	104
6.2.3. Estadística ancillar	105
6.2.4. Estadística suficiente, ancillar y completa	106
7. Estimación puntual	108
7.1. Introducción	108
7.2. Métodos para encontrar estimadores	108
7.2.1. Métodos de momentos	108
7.2.2. Estimadores de máxima verosimilitud	110

7.3.	Métodos de evaluación de estimadores	112
7.3.1.	Error cuadrático medio	112
7.3.2.	Mejores estimadores insesgados	114
7.3.3.	Suficiencia e insesgabilidad	117
7.3.4.	Consistencia	119
7.4.	Otras consideraciones	121
7.4.1.	Variancia asintótica de los estimadores de máxima verosimilitud	121
7.4.2.	Aproximación por series de Taylor	121
8.	Prueba de hipótesis	123
8.1.	Introducción	123
8.2.	Métodos para encontrar pruebas	124
8.2.1.	Prueba de razón de verosimilitud	124
8.3.	Métodos para evaluar pruebas	125
8.3.1.	Probabilidades de error y potencia de prueba	125
8.4.	Pruebas óptimas para hipótesis compuestas	126
8.4.1.	Familias exponenciales	126
8.4.2.	Familias con la propiedad de razón de verosimilitud monótona	127
8.4.3.	Distribución asintótica de la prueba de razón de verosimilitud	128
9.	Estimación por intervalos	130
9.1.	Introducción	130
9.2.	Métodos para encontrar estimadores de intervalos	131
9.2.1.	Invirtiendo una prueba estadística	131
9.2.2.	Cantidades pivotaes	133
9.2.3.	Garantizado un intervalo	134
9.3.	Métodos de evaluación de estimadores por intervalos	136
9.3.1.	Tamaño y probabilidad de cobertura	136
9.4.	Otras consideraciones	137
9.4.1.	Intervalos aproximados por máxima verosimilitud	137
9.4.2.	Otros intervalos aproximados	138

Capítulo 1

Teoría de la probabilidad

1.1. Teoría de conjuntos

Definición 1.1.1 El conjunto S de todos los posibles resultados de un experimento aleatorio es llamado el *espacio muestral*. Un espacio muestral puede ser numerable o no numerable.

Si los elementos del espacio muestral pueden ser puestos en correspondencia 1-1 con algún subconjunto de los números enteros entonces se dice que es numerable, de otro modo el espacio muestral es no numerable.

Definición 1.1.2 Un *evento* es cualquier colección de posibles resultados de un experimento aleatorio, es decir cualquier subconjunto de S (incluyéndolo).

Sean A y B eventos definidos en S :

$$A \subset B \Leftrightarrow x \in A \Rightarrow x \in B$$

$$A = B \Leftrightarrow A \subset B \text{ y } B \subset A$$

La *unión* de A y B , definida como $A \cup B$, es el evento formado por los elementos de S que pertenecen por lo menos a uno de los eventos.

$$A \cup B = \{x : x \in A \text{ o } x \in B\}$$

La *intersección* de A y B , definida como $A \cap B$, es el evento formado por los elementos de S que pertenecen a ambos eventos.

$$A \cap B = \{x : x \in A \text{ y } x \in B\}$$

El *complemento* de A , definido como A^c , es el evento formado por los elementos de S que no pertenecen a A .

$$A^c = \{x : x \notin A\}$$

Ejemplo 1.1.1 Considere el experimento que consiste en elegir al azar una carta de una baraja. Si se está interesado en la figura obtenida en la carta el espacio muestral es:

$$S = \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$$

Algunos posibles eventos son:

$$A = \{\clubsuit, \diamondsuit\} \text{ y } B = \{\diamondsuit, \heartsuit, \spadesuit\}$$

A partir de estos eventos se pueden formar:

$$A \cup B = \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}, A \cap B = \{\diamondsuit\} \text{ y } A^c = \{\heartsuit, \spadesuit\}$$

Además, notar que $A \cup B = S$ y $(A \cup B)^c = \phi$, que denota el *conjunto vacío*.

Teorema 1.1.1 Sean A , B y C eventos definidos en un espacio muestral S ,

a. Conmutatividad:

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

b. Asociatividad:

$$A \cup (B \cup C) = (A \cup B) \cup C$$

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

c. Leyes distributivas:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

d. Leyes de DeMorgan:

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

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

Las operaciones de unión e intersección pueden ser extendidas hacia colecciones infinitas de eventos. Si A_1, A_2, \dots es una colección infinita de eventos definidos sobre un espacio muestral S , entonces

$$\bigcup_{i=1}^{\infty} A_i = \{x \in S : x \in A_i \text{ para algún } i\}$$

$$\bigcap_{i=1}^{\infty} A_i = \{x \in S : x \in A_i \text{ para todo } i\}$$

Ejemplo 1.1.2 Sea $S = (0, 1]$ y se define $A_i = [(1/i), 1]$. Entonces $\bigcup_{i=1}^{\infty} A_i = (0, 1]$ y $\bigcap_{i=1}^{\infty} A_i = \{1\}$.

También es posible definir uniones e intersecciones sobre una colección no numerable de eventos. Si Γ es un conjunto de índices, entonces

$$\bigcup_{\alpha \in \Gamma} A_{\alpha} = \{x \in S : x \in A_{\alpha} \text{ para algún } \alpha\}$$

$$\bigcap_{\alpha \in \Gamma} A_{\alpha} = \{x \in S : x \in A_{\alpha} \text{ para todo } \alpha\}$$

Ejemplo 1.1.3 Si se toma $\Gamma = \{\text{Todos los números reales positivos}\}$ y $A_a = (0, a]$ entonces $\bigcup_{\alpha \in \Gamma} A_{\alpha} = (0, \infty)$ es una unión no numerable.

Definición 1.1.3 Dos eventos A y B son *disjuntos* (o *mutamente excluyentes*) si $A \cap B = \phi$. Los eventos A_1, A_2, \dots son *disjuntos por pares* si $A_i \cap A_j = \phi$ para todo $i \neq j$.

Ejemplo 1.1.4 La colección $A_i = [i, i + 1)$, $i = 0, 1, \dots$ consiste de eventos disjuntos por pares. Notar también que $\bigcup_{i=0}^{\infty} A_i = [0, \infty)$.

Definición 1.1.4 Si A_1, A_2, \dots son disjuntos por pares y $\bigcup_{i=1}^{\infty} A_i = S$ entonces la colección A_1, A_2, \dots forma una *partición* de S .

Ejemplo 1.1.5 Los eventos $A_i = [i, i + 1)$ forman una partición de $S = [0, \infty)$ para $i = 0, 1, \dots$.

1.2. Fundamentos de la teoría de la probabilidad

1.2.1. Axiomas de la probabilidad

Para cada evento A definido en el espacio muestral S es posible asociarle un número entre cero y uno llamado la probabilidad de A y denotado por $\Pr(A)$.

Definición 1.2.1 Una colección de subconjuntos de S es llamada un *sigma álgebra* (o conjunto de Borel), denotada por \mathcal{B} , si satisface las siguientes propiedades:

1. $\phi \in \mathcal{B}$ (el conjunto vacío es un elemento de \mathcal{B}).
2. Si $A \in \mathcal{B}$ entonces $A^c \in \mathcal{B}$ (\mathcal{B} es cerrado bajo complementos).
3. Si $A_1, A_2, \dots \in \mathcal{B}$ entonces $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$ (\mathcal{B} es cerrado bajo uniones numerables).

Definición 1.2.2 Dado un espacio muestral S y un sigma álgebra asociado \mathcal{B} , una *función de probabilidad* es una función \Pr con dominio en \mathcal{B} que satisface:

1. $\Pr(A) \geq 0$ para todo $A \in \mathcal{B}$.
2. $\Pr(S) = 1$.
3. Si $A_1, A_2, \dots \in \mathcal{B}$ son eventos disjuntos por pares, entonces $\Pr(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$.

1.2.2. Cálculo de probabilidades

Teorema 1.2.1 Si \Pr es una función de probabilidad y A es cualquier evento en \mathcal{B} , entonces:

- a. $\Pr(\phi) = 0$, donde ϕ es el conjunto vacío.
- b. $\Pr(A) \leq 1$.
- c. $\Pr(A^c) = 1 - \Pr(A)$.

Teorema 1.2.2 Si \Pr es una función de probabilidad, A y B son eventos en \mathcal{B} , entonces:

- a. $\Pr(B \cap A^c) = \Pr(B) - \Pr(A \cap B)$.
- b. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.
- c. Si $A \subset B$ entonces $\Pr(A) \leq \Pr(B)$.

Teorema 1.2.3 Si \Pr es una función de probabilidad, entonces:

- a. $\Pr(A) = \sum_{i=1}^{\infty} \Pr(A \cap C_i)$ para cualquier partición C_1, C_2, \dots
- b. $\Pr(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \Pr(A_i)$ para eventos cualesquiera A_1, A_2, \dots (Desigualdad de Boole's).

1.2.3. Conteo

Ejemplo 1.2.1 La Tinka es una modalidad de juego de lotería electrónica que consiste en la extracción de seis bolillas sin reemplazo desde un bolillero cerrado que contiene cuarenta y cinco bolillas numeradas del 1 al 45. Para calcular la probabilidad de ganar en este juego es necesario saber cuantos grupos diferentes de seis números pueden escogerse a partir de los cuarenta y cinco.

Ejemplo 1.2.2 En un torneo de eliminación simple, como el torneo abierto de tenis, los participantes avanzan hacia la final solo si ganan. Si se tienen 16 participantes se podría estar interesados en la secuencia de oponentes que debe enfrentar un participante para llegar a la final del torneo.

Teorema 1.2.4 Si un trabajo consistente en k actividades separadas, la i -ésima operación puede realizarse de n_i formas, $i = 1, 2, \dots, k$, entonces el trabajo completo puede realizarse de $n_1 \times n_2 \times \dots \times n_k$ formas.

Definición 1.2.3 Para un entero positivo n , el factorial de n , denotado por $n!$, es el producto de todos los enteros positivos menores o iguales a n . Es decir:

$$n! = n \times (n-1) \times (n-2) \times \dots \times 3 \times 2 \times 1$$

Además, se define $0! = 1$.

Definición 1.2.4 Para dos enteros no negativos n y r , $n \geq r$, se define el símbolo $\binom{n}{r}$ como:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Para saber el número total de jugadas necesarias para ganar el juego de la Tinka podrían considerarse las siguientes posibilidades:

1. *Sin reemplazo y considerando que el orden es importante.* Usando el teorema 1.2.4 el primer número puede ser elegido de 45 formas, el segundo de 44, etc. Es decir, existen:

$$45 \times 44 \times 43 \times 42 \times 41 \times 40 = 5864443200$$

posibles jugadas.

2. *Con reemplazo y considerando que el orden es importante.* Como cada número puede ser elegido de 45 formas, existen:

$$45 \times 45 \times 45 \times 45 \times 45 \times 45 = 45^6 = 8303765625$$

posibles jugadas.

3. *Sin reemplazo y considerando que el orden no es importante.* Luego de hallar el número de jugadas considerando que el orden es importante hay que dividir el resultado entre las jugadas redundantes. Nuevamente por el teorema 1.2.4 seis números pueden ser dispuestos de $6 \times 5 \times 4 \times 3 \times 2 \times 1$ formas, luego el número total de jugadas es:

$$\frac{45 \times 44 \times 43 \times 42 \times 41 \times 40}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = \frac{45!}{6!39!} = 8145060$$

4. *Con reemplazo y considerando que el orden no es importante.* Para realizar el proceso de conteo en este caso se puede considerar que hay 45 casilleros para los números en los que hay que colocar 6 bolillas, digamos B, tal como se muestra a continuación:

Tabla 1.1: 45 casilleros y 6 bolillas

B		BB	B		...		B	B	
1	2	3	4	5	...	42	43	44	45

El número de jugadas posibles es igual al número de formas en que pueden colocarse las 6 bolillas en los 45 casilleros. El trabajo puede resultar menos complicado si consideramos la disposición de las bolillas y las paredes de las cajas sin tomar en cuenta la de los extremos. Luego debe contarse el número total de arreglos de 46 paredes y 6 bolillas. Se tienen 52 objetos que pueden disponerse de $52!$ formas y para eliminar los ordenamientos redundantes luego hay que dividir entre $6!$ y $46!$ dando un total de:

$$\frac{52!}{6!46!} = 20358520$$

Las cuatro situaciones anteriores se resumen a continuación:

Tabla 1.2: Número de posibles arreglos de tamaño r a partir de n objetos

	Sin reemplazo	Con reemplazo
El orden es importante	$\frac{n!}{(n-r)!}$	n^r
El orden no es importante	$\binom{n}{r}$	$\binom{n+r-1}{r}$

1.2.4. Puntos igualmente probables

Suponga que $S = \{s_1, \dots, s_N\}$ es un espacio muestral finito. Se dice que los puntos en S son igualmente probables si $\Pr(\{s_i\}) = \frac{1}{N}$, para todo punto s_i . Luego, usando (3) de la definición de probabilidad, se tiene que para todo evento A :

$$\Pr(A) = \sum_{s_i \in A} \frac{1}{N} = \frac{\text{Número de elementos en } A}{\text{Número de elementos en } S}$$

1.3. Probabilidad condicional e independencia

Definición 1.3.1 Si A y B son eventos en S y $\Pr(B) > 0$, entonces la *probabilidad condicional de A dado B* , representada por $\Pr(A|B)$, es:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad (1.3.1)$$

Teorema 1.3.1 (Regla de Bayes) Sean A_1, A_2, \dots una partición del espacio muestral S y sea B un evento definido también en S . Entonces, para cada $i = 1, 2, \dots$

$$\Pr(A_i|B) = \frac{\Pr(B|A_i) \Pr(A_i)}{\sum_{j=1}^{\infty} \Pr(B|A_j) \Pr(A_j)}$$

Definición 1.3.2 Dos eventos A y B son *estadísticamente independientes* si:

$$\Pr(A \cap B) = \Pr(A) \Pr(B) \quad (1.3.2)$$

Teorema 1.3.2 Si A y B son eventos independientes, entonces los siguientes pares también lo son:

- a. A y B^c .
- b. A^c y B .
- c. A^c y B^c .

Definición 1.3.3 Una colección de eventos A_1, \dots, A_n son *mutuamente independientes* si para cualquier subcolección A_{i_1}, \dots, A_{i_k} se tiene:

$$\Pr\left(\bigcap_{i=1}^k A_{i_j}\right) = \prod_{j=1}^k \Pr(A_{i_j})$$

1.4. Variables aleatorias

Definición 1.4.1 Una *variable aleatoria* es una función que se define desde un espacio muestral S hacia los números reales.

1.5. Función de distribución acumulada

Definición 1.5.1 La *función de distribución acumulada* de una variable aleatoria X , denotada por $F_X(x)$, se define por:

$$F_X(x) = \Pr(X \leq x), \text{ para todo } x$$

Teorema 1.5.1 La función $F_X(x)$ es una función de distribución acumulada si y solo si se cumplen las tres condiciones siguientes:

- a. $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$.
- b. $F(x)$ es una función no decreciente de x .
- c. $F(x)$ es continua hacia la derecha; esto es, para todo número x_0 , $\lim_{x \downarrow x_0} F(x) = F(x_0)$.

Definición 1.5.2 Una variable aleatoria X es *continua* si $F_X(x)$ es una función continua de x . Una variable aleatoria X es *discreta* si $F_X(x)$ es una función paso de x .

Definición 1.5.3 Las variables aleatorias X y Y son *identicamente distribuidas* si para cada evento $A \in \mathcal{B}$, $\Pr(X \in A) = \Pr(Y \in A)$.

Teorema 1.5.2 Las dos afirmaciones siguientes son equivalentes:

- a. Las variables aleatorias X y Y son identicamente distribuidas.
- b. $F_X(x) = F_Y(x)$, para todo x .

1.6. Función de probabilidad y densidad

Definición 1.6.1 La *función de probabilidad* de una variable aleatoria discreta X esta dada por:

$$f_X(x) = \Pr(X = x), \text{ para todo } x$$

Definición 1.6.2 La *función de densidad*, $f_X(x)$, de una variable aleatoria continua X es la función que satisface:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \text{ para todo } x$$

Teorema 1.6.1 Una función $f_X(x)$ es una función de probabilidad o función de densidad para una variable aleatoria X si y solo si:

- a. $f_X(x) \geq 0$ para todo x .
- b. $\sum_x f_X(x) = 1$ si X es variable aleatoria discreta y $\int_{-\infty}^{\infty} f_X(x) dx = 1$ si X es variable aleatoria continua.

Capítulo 2

Transformaciones y esperanza

2.1. Transformaciones para variables aleatorias

Si X es una variable aleatoria con función de distribución acumulada $F_X(x)$, entonces cualquier función de X es también una variable aleatoria. Si se define $Y = g(X)$ es posible describir el comportamiento probabilístico de Y en términos de X .

Formalmente, $y = g(x)$ define un mapa desde el espacio muestral de X , \mathcal{X} , al espacio muestral de Y , \mathcal{Y} . Es decir:

$$g(x) : \mathcal{X} \longrightarrow \mathcal{Y}$$

Se asocia a g un mapa inverso, denotado por g^{-1} , definido por:

$$g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\} \quad (2.1.1)$$

$g^{-1}(A)$ es el conjunto de puntos en \mathcal{X} tal que $g(x)$ pertenece al conjunto A . Si el conjunto A es unitario, digamos $A = \{y\}$, entonces:

$$g^{-1}(\{y\}) = \{x \in \mathcal{X} : g(x) = y\}$$

En este caso podría escribirse $g^{-1}(y)$ en lugar de $g^{-1}(\{y\})$. Si existe un único x tal que $g(x) = y$, entonces $g^{-1}(y)$ es el conjunto unitario $\{x\}$ y además $g^{-1}(y) = x$.

Si se define $Y = g(X)$, entonces para cualquier conjunto $A \subset \mathcal{Y}$:

$$\begin{aligned}
\Pr(Y \in A) &= \Pr(g(X) \in A) \\
&= \Pr(\{x \in \mathcal{X} : g(x) \in A\}) \\
&= \Pr(X \in g^{-1}(A))
\end{aligned} \tag{2.1.2}$$

2.1.1. Caso discreto

Si X es una variable aleatoria discreta entonces \mathcal{X} es numerable. El espacio muestral para $Y = g(X)$ es $\mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\}$ el cual también es un conjunto numerable. Usando la ecuación 2.1.2 la función de probabilidad de Y es:

$$f_Y(y) = \Pr(Y = y) = \sum_{x \in g^{-1}(y)} \Pr(X = x) = \sum_{x \in g^{-1}(y)} f_X(x)$$

Ejemplo 2.1.1 Sea $X \sim \mathcal{BI}(n, p)$ cuya función de probabilidad es:

$$f_X(x) = \Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \tag{2.1.3}$$

para $x = 0, 1, \dots, n$ y $0 \leq p \leq 1$. Considere la variable aleatoria $Y = g(X) = n - X$, entonces $\mathcal{X} = \{0, 1, \dots, n\}$ y $\mathcal{Y} = \{0, 1, \dots, n\}$. Luego $g^{-1}(y)$ es el conjunto unitario $x = n - y$, tal que:

$$\begin{aligned}
f_Y(y) &= \sum_{x \in g^{-1}(y)} f_X(x) \\
&= f_X(n - y) \\
&= \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)} \\
&= \binom{n}{y} (1-p)^y p^{n-y}
\end{aligned}$$

es decir, $Y \sim \mathcal{BI}(n, 1-p)$.

2.1.2. Caso continuo

Si X y $Y = g(X)$ son variables aleatorias continuas es posible encontrar fórmulas para obtener la función de distribución acumulada y la función de densidad de Y en términos de F_X , f_X y la función g .

La función de distribución acumulada de $Y = g(X)$ es:

$$\begin{aligned}
 F_Y(y) &= \Pr(Y \leq y) \\
 &= \Pr(g(X) \leq y) \\
 &= \Pr(\{x \in \mathcal{X} : g(x) \leq y\}) \\
 &= \int_{\{x \in \mathcal{X} : g(x) \leq y\}} f_X(x) dx
 \end{aligned} \tag{2.1.4}$$

Suelen existir algunas dificultades en identificar $\{x \in \mathcal{X} : g(x) \leq y\}$ y llevar a cabo el proceso de integración de $f_X(x)$ sobre esta región. Suele ser conveniente utilizar:

$$\mathcal{X} = \{x : f_X(x) > 0\} \text{ y } \mathcal{Y} = \{y : y = g(x) \text{ para algún } x \in \mathcal{X}\} \tag{2.1.5}$$

La función de densidad de la variable aleatoria X es positiva solo sobre \mathcal{X} llamado el *conjunto soporte* de la distribución. Esta terminología también es aplicable a una función de probabilidad.

Es sencillo trabajar con funciones $g(x)$ que son *monótonas*, es decir, aquellas que satisfacen alguna de las siguientes relaciones:

$$u > v \Rightarrow g(u) > g(v) \text{ (creciente) } \quad \text{o} \quad u < v \Rightarrow g(u) > g(v) \text{ (decreciente)}$$

Si la transformación $x \rightarrow g(x)$ es monótona entonces es *uno a uno* y *sobreyectiva*. La transformación es uno a uno si cada x permite obtener un único y y cada y se obtiene con a lo más un x . La transformación es sobreyectiva si para cada $y \in \mathcal{Y}$, definido en 2.1.5, existe un $x \in \mathcal{X}$ tal que $g(x) = y$. Si g es monótona, entonces g^{-1} es unitario; es decir, $g^{-1}(y) = x$ si y solo si $y = g(x)$.

Si g es creciente, esto implica que:

$$\begin{aligned}
 \{x \in \mathcal{X} : g(x) \leq y\} &= \{x \in \mathcal{X} : g^{-1}(g(x)) \leq g^{-1}(y)\} \\
 &= \{x \in \mathcal{X} : x \leq g^{-1}(y)\}
 \end{aligned} \tag{2.1.6}$$

y usando 2.1.4, se tiene que:

$$F_Y(y) = \int_{\{x \in \mathcal{X} : x \leq g^{-1}(y)\}} f_X(x) dx = \int_{-\infty}^{g^{-1}(y)} f_X(x) dx = F_X(g^{-1}(y))$$

Si g es decreciente, entonces:

$$\begin{aligned}\{x \in \mathcal{X} : g(x) \leq y\} &= \{x \in \mathcal{X} : g^{-1}(g(x)) \geq g^{-1}(y)\} \\ &= \{x \in \mathcal{X} : x \geq g^{-1}(y)\}\end{aligned}\quad (2.1.7)$$

se tiene que:

$$F_Y(y) = \int_{g^{-1}(y)}^{\infty} f_X(x) dx = 1 - F_X(g^{-1}(y))$$

Se resumen los resultados anteriores en el siguiente teorema.

Teorema 2.1.1 Sea X una variable aleatoria con función de distribución acumulada $F_X(x)$. Se define $Y = g(X)$ y los espacios muestrales \mathcal{X} y \mathcal{Y} según 2.1.5.

- a. Si g es una función creciente sobre \mathcal{X} , entonces $F_Y(y) = F_X(g^{-1}(y))$ para $y \in \mathcal{Y}$.
- b. Si g es una función decreciente sobre \mathcal{X} y X es una variable aleatoria continua, entonces $F_Y(y) = 1 - F_X(g^{-1}(y))$ para $y \in \mathcal{Y}$.

Ejemplo 2.1.2 Suponga que X tiene función de densidad $f(x) = 1$ si $0 < x < 1$, luego $F_X(x) = x$. Si se realiza la transformación $Y = g(X) = -\log X$, entonces $g(x)$ es una función decreciente y $\mathcal{Y} = (0, \infty)$. Para $y > 0$, $y = -\log x$ implica $x = g^{-1}(y) = e^{-y}$, luego:

$$F_Y(y) = 1 - F_X(g^{-1}(y)) = 1 - F_X(e^{-y}) = 1 - e^{-y}$$

además $F_Y(y) = 0$ para $y \leq 0$.

La función de densidad de Y puede obtenerse derivando su función de distribución acumulada. La expresión resultante se presenta en el siguiente teorema.

Teorema 2.1.2 Sea X con función de densidad $f_X(x)$ y sea $Y = g(X)$, donde g es una función monótona. Sean \mathcal{X} y \mathcal{Y} tal como se definieron en 2.1.5. Suponga que $f_X(x)$ es continua sobre \mathcal{X} y que $g^{-1}(y)$ tiene una derivada continua sobre \mathcal{Y} . Entonces la función de densidad de Y es:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{de otro modo} \end{cases} \quad (2.1.8)$$

Demostración: Usando el teorema 2.1.1 y la regla de la cadena se tiene:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) & \text{si } g \text{ es creciente} \\ -f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) & \text{si } g \text{ es decreciente} \end{cases}$$

Ejemplo 2.1.3 Sea $f_X(x)$ la función de densidad *gamma*:

$$f_X(x) = \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta}, \quad 0 < x < \infty$$

donde $\beta > 0$ y n es un entero positivo. Suponga que se desea encontrar la función de densidad de $g(X) = 1/X$. Notar que $\mathcal{X} = \mathcal{Y} = (0, \infty)$. Si $y = g(x)$, entonces $g^{-1}(y) = 1/y$ y $\frac{d}{dy} g^{-1}(y) = -1/y^2$. Aplicando el teorema anterior:

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= \frac{1}{(n-1)!\beta^n} \left(\frac{1}{y} \right)^{n-1} e^{-1/(\beta y)} \frac{1}{y^2} \\ &= \frac{1}{(n-1)!\beta^n} \left(\frac{1}{y} \right)^{n+1} e^{-1/(\beta y)} \end{aligned}$$

se obtiene un caso especial de una función de densidad conocida como *gamma inversa*.

En muchas aplicaciones la función g podría no ser creciente ni decreciente, por consiguiente no podrían aplicarse los resultados anteriores. Sin embargo es común el caso en el que la función g es monótona sobre ciertos subintervalos, los que permiten obtener una expresión para $Y = g(X)$.

Ejemplo 2.1.4 Suponga que X es una variable aleatoria continua. La función de distribución acumulada de $Y = X^2$, para $y > 0$, es:

$$F_Y(y) = \Pr(Y \leq y) = \Pr(X^2 \leq y) = \Pr(-\sqrt{y} \leq X \leq \sqrt{y})$$

Como X es variable aleatoria continua se tiene:

$$\begin{aligned} F_Y(y) &= \Pr(-\sqrt{y} < X \leq \sqrt{y}) \\ &= \Pr(X \leq \sqrt{y}) - \Pr(X \leq -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

La función de densidad de Y puede obtenerse derivando su función de distribución acumulada:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} [F_X(\sqrt{y}) - F_X(-\sqrt{y})] \end{aligned}$$

y usando la regla de la cadena para derivar $F_X(\sqrt{y})$ y $F_X(-\sqrt{y})$ se tiene:

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \quad (2.1.9)$$

Notar que la función de densidad anterior esta expresada como la suma de dos componentes sobre los intervalos donde $g(x) = x^2$ es monótona.

Teorema 2.1.3 Sea X con función de densidad $f_X(x)$, $Y = g(X)$ y el espacio muestral \mathcal{X} definido en 2.1.5. Suponga que existe una partición A_0, A_1, \dots, A_k de \mathcal{X} tal que $\Pr(X \in A_0) = 0$ y $f_X(x)$ es continua sobre cada A_i . Suponga además que existen funciones $g_1(x), \dots, g_k(x)$ definidas sobre A_1, \dots, A_k respectivamente, que satisfacen:

- a. $g(x) = g_i(x)$ para $x \in A_i$,
- b. $g_i(x)$ es monótona sobre A_i ,
- c. El conjunto $\mathcal{Y} = \{y : y = g_i(x) \text{ para algún } x \in A_i\}$ es el mismo para cada $i = 1, \dots, k$.
- d. $g_i^{-1}(y)$ tiene una derivada continua en \mathcal{Y} , para cada $i = 1, \dots, k$.

Entonces:

$$f_Y(y) = \begin{cases} \sum_{i=1}^k f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{de otro modo} \end{cases}$$

Es importante notar que cada $g_i(x)$ es una transformación uno a uno desde A_i hacia \mathcal{Y} . Además, $g_i^{-1}(y)$ es una función uno a uno desde \mathcal{Y} hacia A_i , tal que, para $y \in \mathcal{Y}$, $g_i^{-1}(y)$ permite obtener un único $x = g_i^{-1}(y) \in A_i$ para el cual $g_i(x) = y$.

Ejemplo 2.1.5 Sea X con *distribución normal estándar*,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty$$

Si $Y = X^2$, la función $g(x) = x^2$ es monótona sobre $(-\infty, 0)$ y $(0, \infty)$ con lo que $\mathcal{Y} = (0, \infty)$. Aplicando el teorema 2.1.3 se tiene $A_0 = \{0\}$ y

$$\begin{aligned} A_1 &= (-\infty, 0) & g_1(x) &= x^2 & g_1^{-1}(y) &= -\sqrt{y} \\ A_2 &= (0, \infty) & g_2(x) &= x^2 & g_2^{-1}(y) &= \sqrt{y} \end{aligned}$$

La función de densidad de Y es:

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} \left| -\frac{1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} \left| \frac{1}{2\sqrt{y}} \right| \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-y/2}, \quad 0 < y < \infty \end{aligned}$$

la cual es conocida como la distribución *chi cuadrado* con 1 grado de libertad.

Teorema 2.1.4 Sea X cuya función de distribución acumulada, $F_X(x)$, es continua. Si se define la variable aleatoria $Y = F_X(x)$, entonces Y tiene distribución uniforme en el intervalo $(0, 1)$.

Demostración: Si $Y = F_X(x)$ entonces $0 < y < 1$,

$$\begin{aligned} \Pr(Y \leq y) &= \Pr(F_X(X) \leq y) \\ &= \Pr(F_X^{-1}[F_X(X)] \leq F_X^{-1}(y)) \\ &= \Pr(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) \\ &= y \end{aligned}$$

2.2. Valores esperados

Definición 2.2.1 El *valor esperado* o *media* de una variable aleatoria $g(X)$, denotado por $E[g(X)]$, es:

$$E[g(X)] = \begin{cases} \sum_{x \in \mathcal{X}} g(x) f_X(x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{si } X \text{ es continua} \end{cases} \quad (2.2.1)$$

siempre que la integral o suma exista. Si $E[|g(X)|] = \infty$ se dice que $E[g(X)]$ no existe.

Ejemplo 2.2.1 Suponga que $X \sim \mathcal{E}(\lambda)$, entonces su función de densidad es:

$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}$$

para $0 \leq x < \infty$ y $\lambda > 0$. Luego:

$$E[X] = \int_0^{\infty} x \frac{1}{\lambda} e^{-x/\lambda} dx = \lambda$$

Ejemplo 2.2.2 Si $X \sim \mathcal{BI}(n, p)$, entonces:

$$E[X] = \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^n x \binom{n}{x} p^x (1-p)^{n-x}$$

Usando la identidad $x \binom{n}{x} = n \binom{n-1}{x-1}$ se tiene:

$$\begin{aligned} E[X] &= \sum_{x=1}^n n \binom{n-1}{x-1} p^x (1-p)^{n-x} \\ &= \sum_{y=0}^{n-1} n \binom{n-1}{y} p^{y+1} (1-p)^{n-(y+1)} \\ &= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} \\ &= np \end{aligned}$$

Ejemplo 2.2.3 Un ejemplo clásico de una variable aleatoria cuyo valor esperado no existe corresponde a la *distribución de Cauchy* cuya función de densidad es:

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad -\infty < x < \infty$$

Es fácil verificar que $\int_{-\infty}^{\infty} f_X(x) dx = 1$, pero con respecto a su valor esperado:

$$E[|X|] = \int_{-\infty}^{\infty} \frac{|x|}{\pi} \frac{1}{1+x^2} dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx$$

Para cualquier número positivo M ,

$$\int_0^M \frac{x}{1+x^2} dx = \frac{1}{2} \log(1+x^2) \Big|_0^M = \frac{1}{2} \log(1+M^2)$$

Luego,

$$E[|X|] = \lim_{M \rightarrow \infty} \frac{2}{\pi} \int_0^M \frac{x}{1+x^2} dx = \frac{1}{\pi} \lim_{M \rightarrow \infty} \log(1+M^2) = \infty$$

y por consiguiente el $E[X]$ de esta distribución no existe.

Teorema 2.2.1 Sea X una variable aleatoria y sean a , b y c constantes. Entonces para funciones cualesquiera $g_1(x)$ y $g_2(x)$ cuyo valor esperado exista,

- a. $E[ag_1(X) + bg_2(X) + c] = aE[g_1(X)] + bE[g_2(X)] + c$.
- b. Si $g_1(x) \geq 0$ para todo x , entonces $E[g_1(X)] \geq 0$.
- c. Si $g_1(x) \geq g_2(x)$ para todo x , entonces $E[g_1(X)] \geq E[g_2(X)]$.
- d. Si $a \leq g_1(x) \leq b$ para todo x , entonces $a \leq E[g_1(X)] \leq b$.

Ejemplo 2.2.4 Suponga que se mide la distancia entre una variable aleatoria X y una constante b mediante $(X - b)^2$. Mientras más cerca esté b de X más pequeña será dicha cantidad. El objetivo es determinar el valor de b que minimize $E[(X - b)^2]$.

$$\begin{aligned} E[(X - b)^2] &= E[(X - E[X] + E[X] - b)^2] \\ &= E[((X - E[X]) + (E[X] - b))^2] \\ &= E[(X - E[X])^2] + E[(E[X] - b)^2] \end{aligned}$$

ya que $E[(X - E[X])(E[X] - b)] = 0$. Además $(E[X] - b)$ es una constante. Luego:

$$E[(X - b)^2] = E[(X - E[X])^2] + (E[X] - b)^2$$

Como no se tiene control sobre el primer término del lado derecho y el segundo término puede ser mayor o igual a 0, el menor valor se obtiene cuando $b = E[X]$. Entonces:

$$\min_b E[(X - b)^2] = E[(X - E[X])^2]$$

2.3. Momentos y función generatriz de momentos

Definición 2.3.1 Para cada entero n , el n -ésimo momento de X , μ'_n , es:

$$\mu'_n = E[X^n]$$

El n -ésimo momento central de X , μ_n , es:

$$\mu_n = E[(X - \mu)^n]$$

donde $\mu = \mu'_1 = E[X]$.

Definición 2.3.2 La *varianza* de una variable aleatoria X es su segundo momento central, $\text{Var}(X) = E[(X - \mu)^2]$. La raíz cuadrada positiva de la varianza es conocida como *desviación estándar*.

Ejemplo 2.3.1 Si $X \sim \mathcal{E}(\lambda)$, entonces:

$$\text{Var}(X) = E[(X - \mu)^2] = \int_0^{\infty} (x - \lambda)^2 \frac{1}{\lambda} e^{-x/\lambda} dx = \lambda^2$$

Teorema 2.3.1 Si X es una variable aleatoria con varianza finita, entonces para constantes cualesquiera a y b :

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Demostración: Usando la definición de varianza:

$$\begin{aligned} \text{Var}(aX + b) &= E[((aX + b) - E[(aX + b)])^2] \\ &= E[(aX - aE[X])^2] \\ &= a^2 E[(X - E[X])^2] \\ &= a^2 \text{Var}(X) \end{aligned}$$

La siguiente forma de calcular la varianza es bastante útil:

$$\text{Var}(X) = E[X^2] - E^2[X] \quad (2.3.1)$$

Ejemplo 2.3.2 Si $X \sim \mathcal{BI}(n, p)$, entonces:

$$\mathbb{E}[X^2] = \sum_{x=0}^n x^2 \binom{n}{x} p^x (1-p)^{n-x}$$

pero:

$$x^2 \binom{n}{x} = x \frac{n!}{(x-1)!(n-x)!} = xn \binom{n-1}{x-1}$$

luego,

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{x=1}^n xn \binom{n-1}{x-1} p^x (1-p)^{n-x} \\ &= n \sum_{y=0}^{n-1} (y+1) \binom{n-1}{y} p^{y+1} (1-p)^{n-1-y} \\ &= np(n-1)p + np \end{aligned}$$

Finalmente:

$$\text{Var}[X] = n^2 p^2 - np^2 + np - (np)^2 = np(1-p)$$

Definición 2.3.3 Sea X una variable aleatoria. La *función generatriz de momentos* de X , denotada por $M_X(t)$, es:

$$M_X(t) = \mathbb{E}[e^{tX}]$$

sujeto a que el valor esperado exista para t en alguna vecindad de 0. Es decir, existe $h > 0$ tal que, para todo t en $-h < t < h$, $\mathbb{E}[e^{tX}]$ existe.

Más explícitamente:

$$M_X(t) = \begin{cases} \sum_x e^{tx} f_X(x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x) dx & \text{si } X \text{ es continua} \end{cases}$$

Teorema 2.3.2 Si X tiene función generatriz de momentos $M_X(t)$ entonces:

$$\mathbb{E}[X] = M_X^{(1)}(0)$$

donde $M_X^{(n)}(0) = \left. \frac{d^n}{dt^n} M_X(t) \right|_{t=0}$.

Prueba: Asumiendo que es posible intercambiar la derivada con la integral, se tiene:

$$\begin{aligned}
 \frac{d}{dt}M_X(t) &= \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \\
 &= \int_{-\infty}^{\infty} \left(\frac{d}{dt} e^{tx} \right) f_X(x) dx \\
 &= \int_{-\infty}^{\infty} x e^{tx} f_X(x) dx \\
 &= E[X e^{tX}]
 \end{aligned}$$

luego $\frac{d}{dt}M_X(t)\big|_{t=0} = E[X e^{tX}]\big|_{t=0} = E[X]$. Trabajando de manera análoga, se puede establecer que:

$$\frac{d^n}{dt^n}M_X(t)\big|_{t=0} = E[X^n e^{tX}]\big|_{t=0} = E[X^n]$$

Ejemplo 2.3.3 En el ejemplo 2.1.3 se usó un caso especial de la función de densidad gamma:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad 0 < x < \infty, \alpha > 0, \beta > 0$$

donde $\Gamma(\alpha)$ denota la función gamma cuyas propiedades se mencionan en la sección 3.2. La función generatriz de momentos de la distribución gamma esta dada por:

$$\begin{aligned}
 M_X(t) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty e^{tx} x^{\alpha-1} e^{-x/\beta} dx \\
 &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x((1/\beta)-t)} dx \\
 &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x/(\frac{\beta}{1-\beta t})} dx \\
 &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \Gamma(\alpha) \left(\frac{\beta}{1-\beta t} \right)^\alpha \\
 &= \left(\frac{1}{1-\beta t} \right)^\alpha
 \end{aligned}$$

y existe solo si $t < 1/\beta$. La media de la distribución gamma es:

$$E[X] = \frac{d}{dt} M_X(t) \Big|_{t=0} = \frac{\alpha\beta}{(1-\beta t)^{\alpha+1}} \Big|_{t=0} = \alpha\beta$$

Los otros momentos pueden calcularse de forma similar.

Ejemplo 2.3.4 Si $X \sim \mathcal{BI}(n, p)$, entonces:

$$\begin{aligned} M_X(t) &= \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= [pe^t + (1-p)]^n \end{aligned}$$

recordando que $\sum_{x=0}^n \binom{n}{x} u^x v^{n-x} = (u+v)^n$.

Si la función generatriz de momentos existe, entonces caracteriza un conjunto infinito de momentos. La pregunta natural es si la condición anterior determina una función de distribución acumulada única.

Ejemplo 2.3.5 Considere las siguientes funciones de densidad dadas por:

$$\begin{aligned} f_1(x) &= \frac{1}{\sqrt{2\pi x}} e^{-(\log x)^2/2}, \quad 0 \leq x < \infty \\ f_2(x) &= f_1(x) [1 + \sin(2\pi \log x)], \quad 0 \leq x < \infty \end{aligned}$$

La función de densidad $f_1(x)$ es un caso especial de la *función de densidad lognormal*. Se puede probar que si $X_1 \sim f_1(x)$ entonces:

$$E[X_1^r] = e^{r^2/2}, \quad r = 0, 1, 2, \dots$$

es decir, X_1 tiene todos sus momentos. Ahora si $X_2 \sim f_2(x)$ se tiene:

$$\begin{aligned} E[X_2^r] &= \int_0^\infty x^r f_1(x) [1 + \sin(2\pi \log x)] dx \\ &= E[X_1^r] + \int_0^\infty x^r f_1(x) \sin(2\pi \log x) dx \end{aligned}$$

La transformación $y = \log(x) - r$ muestra que la última integral sobre una función impar en $(-\infty, \infty)$ es igual a cero para $r = 0, 1, 2, \dots$. Es decir, que aunque X_1 y X_2 tienen diferentes funciones de densidad sus momentos son iguales para todo r . Las dos funciones de densidad son graficadas en la figura 2.1.

Teorema 2.3.3 Sean $F_X(x)$ y $F_Y(y)$ dos funciones de distribución acumulada tal que todos sus momentos existen.

- a. Si F_X y F_Y tienen soporte acotado, entonces $F_X(u) = F_Y(u)$ para todo u si y solo si $E[X^r] = E[Y^r]$ para todo entero $r = 0, 1, 2, \dots$.
- b. Si las funciones generatrices de momentos existen y $M_X(t) = M_Y(t)$ para todo t en alguna vecindad de cero, entonces $F_X(u) = F_Y(u)$ para todo u .

Teorema 2.3.4 Suponga $\{X_i, i = 1, 2, \dots\}$ es una secuencia de variables aleatorias cuya función generatriz de momentos es $M_{X_i}(t)$. Además:

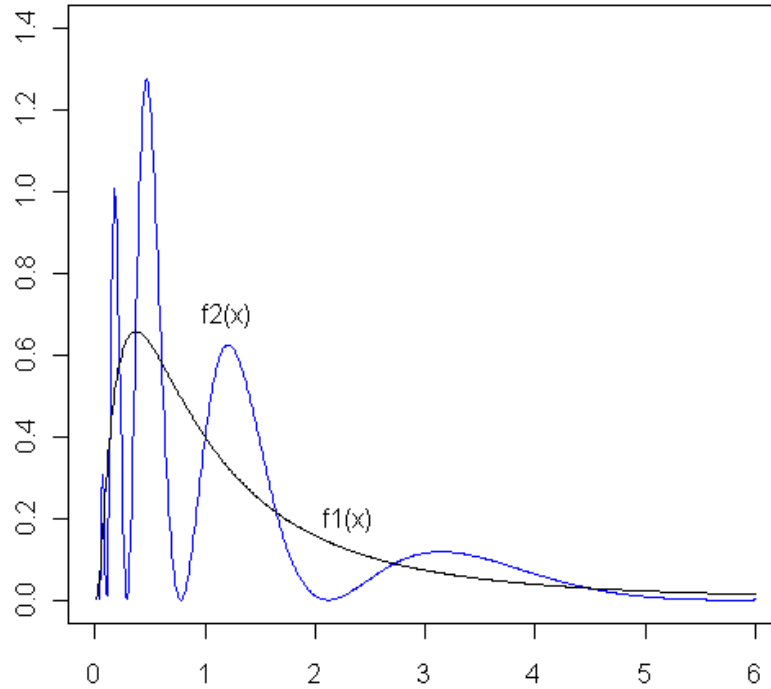
$$\lim_{i \rightarrow \infty} M_{X_i}(t) = M_X(t) \quad \text{para todo } t \text{ en una vecindad de cero}$$

donde $M_X(t)$ es una función generatriz de momentos. Entonces existe una única función de distribución acumulada F_X cuyos momentos estan definidos por $M_X(t)$ y, para todo x donde $F_X(x)$ es continua, se tiene:

$$\lim_{i \rightarrow \infty} F_{X_i}(x) = F_X(x)$$

Es decir, la convergencia, para $|t| < h$, de funciones generatrices de momentos en una función generatriz de momentos implica convergencia de funciones de distribución acumulada.

Figura 2.1: Dos funciones de densidad con los mismos momentos



Ejemplo 2.3.6 Una aproximación usada en cursos elementales de estadística permite aproximar las probabilidades binomiales usando la distribución de Poisson. Esta aproximación es válida *cuando n es grande y np es pequeño*.

La función de probabilidad de Poisson es:

$$\Pr(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, 2, \dots$$

donde λ es una constante positiva. La aproximación es tal que si X tiene distribución binomial(n, p) y Y tiene distribución de Poisson(λ) con $\lambda = np$, entonces:

$$\Pr(X = x) \approx \Pr(Y = x)$$

Recordar que:

$$M_X(t) = [pe^t + (1-p)]^n$$

es la función generatriz de momentos de la distribución binomial. Para la distribución de Poisson se puede demostrar que su función generatriz de momentos es:

$$M_Y(t) = e^{\lambda(e^t-1)}$$

Como $\lambda = np$, entonces:

$$\begin{aligned} M_X(t) &= [1 + p(e^t - 1)]^n \\ &= \left[1 + \frac{1}{n}(e^t - 1)\lambda\right]^n \end{aligned}$$

Lema 2.3.1 Sean a_1, a_2, \dots una secuencia de números que convergen hacia a , es decir $\lim_{n \rightarrow \infty} a_n = a$, entonces:

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$$

Demostración: La demostración de este lema puede encontrarse en los textos de cálculo.

Luego, si se toma $a_n = \lambda(e^t - 1) = a$ entonces:

$$\lim_{n \rightarrow \infty} M_X(t) = e^{\lambda(e^t-1)} = M_Y(t)$$

es la función generatriz de momentos de la distribución de Poisson.

Teorema 2.3.5 Sean a y b constantes, la función generatriz de momentos de la variable aleatoria $aX + b$ está dada por:

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

Prueba: Por definición:

$$\begin{aligned} M_{aX+b}(t) &= \mathbb{E} \left[e^{(aX+b)t} \right] \\ &= \mathbb{E} \left[e^{(aX)t} e^{bt} \right] \\ &= e^{bt} \mathbb{E} \left[e^{(aX)t} \right] \\ &= e^{bt} M_X(at) \end{aligned}$$

2.4. Derivación bajo el signo de la integral

Teorema 2.4.1 (Regla de Leibnit'z) Si $f(x, \theta)$, $a(\theta)$ y $b(\theta)$ son diferenciables con respecto a θ , entonces:

$$\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x, \theta) dx = f(b(\theta), \theta) \frac{d}{d\theta} b(\theta) - f(a(\theta), \theta) \frac{d}{d\theta} a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial \theta} f(x, \theta) dx$$

Notar que si $a(\theta)$ y $b(\theta)$ son constantes, se tiene un caso especial de la regla anterior:

$$\frac{d}{d\theta} \int_a^b f(x, \theta) dx = \int_a^b \frac{\partial}{\partial \theta} f(x, \theta) dx$$

El rango en la integral anterior es finito, en caso contrario podrían surgir complicaciones.

Teorema 2.4.2 Suponga que la función $h(x, y)$ es continua en y_0 para todo x , y existe una función $g(x)$ que satisface:

- a. $|h(x, y)| \leq g(x)$ para todo x y y ,
- b. $\int_{-\infty}^{\infty} g(x) dx < \infty$,

entonces:

$$\lim_{y \rightarrow y_0} \int_{-\infty}^{\infty} h(x, y) dx = \int_{-\infty}^{\infty} \lim_{y \rightarrow y_0} h(x, y) dx$$

Es posible aplicar el teorema anterior para el caso en que $h(x, y)$ se identifique con la diferencia $(f(x, \theta + \delta) - f(x, \theta))/\delta$.

Teorema 2.4.3 Suponga que $f(x, \theta)$ es diferenciable en $\theta = \theta_0$, es decir,

$$\lim_{\delta \rightarrow 0} \frac{f(x, \theta + \delta) - f(x, \theta)}{\delta} = \frac{\partial}{\partial \theta} f(x, \theta) \Big|_{\theta = \theta_0}$$

para todo x , y que existen una función $g(x, \theta_0)$ y una constante $\delta_0 > 0$ tales que:

- a. $\left| \frac{f(x, \theta + \delta) - f(x, \theta)}{\delta} \right| \leq g(x, \theta_0)$ para todo x y $|\delta| \leq \delta_0$,
- b. $\int_{-\infty}^{\infty} g(x, \theta_0) dx < \infty$,

entonces:

$$\frac{d}{d\theta} \int_{-\infty}^{\infty} f(x, \theta) dx \Big|_{\theta=\theta_0} = \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \theta} f(x, \theta) \Big|_{\theta=\theta_0} \right] dx \quad (2.4.1)$$

A menudo no se hace hincapié en la distinción entre θ y θ_0 , luego 2.4.1 se escribe como:

$$\frac{d}{d\theta} \int_{-\infty}^{\infty} f(x, \theta) dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} f(x, \theta) dx \quad (2.4.2)$$

Corolario 2.4.1 Suponga $f(x, \theta)$ es diferenciable en θ y existe una función $g(x, \theta)$ tal que:

$$\left| \frac{\partial}{\partial \theta} f(x, \theta) \Big|_{\theta=\theta'} \right| \leq g(x, \theta) \text{ para todo } \theta' \text{ tal que } |\theta' - \theta| \leq \delta_0$$

y $\int_{-\infty}^{\infty} g(x, \theta) dx < \infty$, entonces se cumple 2.4.2.

Ejemplo 2.4.1 Sea $X \sim \mathcal{E}(\lambda)$. Suponga que se desea calcular:

$$\frac{d}{d\lambda} \mathbb{E}[X^n] = \frac{d}{d\lambda} \int_0^{\infty} x^n \left(\frac{1}{\lambda} \right) e^{-x/\lambda} dx$$

para $n > 0$ número entero. Si se desea intercambiar la derivada con la integral, se tiene:

$$\begin{aligned} \frac{d}{d\lambda} \mathbb{E}[X^n] &= \int_0^{\infty} \frac{d}{d\lambda} x^n \left(\frac{1}{\lambda} \right) e^{-x/\lambda} dx \\ &= \int_0^{\infty} \frac{x^n}{\lambda^2} \left(\frac{x}{\lambda} - 1 \right) e^{-x/\lambda} dx \\ &= \frac{1}{\lambda^2} \mathbb{E}[X^{n+1}] - \frac{1}{\lambda} \mathbb{E}[X^n] \end{aligned}$$

Para justificar el intercambio anterior, se acota la derivada de $x^n \left(\frac{1}{\lambda} \right) e^{-x/\lambda}$,

$$\left| \frac{\partial}{\partial \lambda} \left(\frac{x^n e^{-x/\lambda}}{\lambda} \right) \right| = \frac{x^n e^{-x/\lambda}}{\lambda^2} \left| \left(\frac{x}{\lambda} - 1 \right) \right| \leq \frac{x^n e^{-x/\lambda}}{\lambda^2} \left(\frac{x}{\lambda} + 1 \right)$$

desde que $x/\lambda > 0$. Para alguna constante δ_0 que satisface $0 < \delta_0 < \lambda$, se toma:

$$g(x, \lambda) = \frac{x^n e^{-x/(\lambda+\delta_0)}}{(\lambda - \delta_0)^2} \left(\frac{x}{\lambda - \delta_0} + 1 \right)$$

entonces:

$$\left| \frac{\partial}{\partial \lambda} \left(\frac{x^n e^{-x/\lambda}}{\lambda} \right) \Big|_{\lambda=\lambda'} \right| \leq g(x, \lambda) \text{ para todo } \lambda' \text{ tal que } |\lambda' - \lambda| \leq \delta_0$$

Como la distribución exponencial tiene todos sus momentos, se cumple $\int_{-\infty}^{\infty} g(x, \theta) dx < \infty$ siempre que $\lambda - \delta_0 > 0$, de manera que se justifica el cambio de la derivada con la integral.

La propiedad ilustrada para la distribución exponencial se cumple para una gran clase de densidades de la sección 3.2.

Ejemplo 2.4.2 Sea $X \sim \mathcal{N}(\mu, \sigma^2 = 1)$ y considere su función generatriz de momentos:

$$M_X(t) = E[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-(x-\mu)^2/2} dx$$

entonces:

$$\frac{d}{dt} M_X(t) = \frac{d}{dt} E[e^{tX}] = E \left[\frac{\partial}{\partial t} e^{tX} \right] = E[X e^{tX}]$$

Ejemplo 2.4.3 Sea $X \sim \mathcal{GE}(\theta)$ tal que:

$$\Pr(X = x) = \theta(1 - \theta)^x \quad x = 0, 1, \dots \quad 0 < \theta < 1$$

luego:

$$\begin{aligned} \frac{d}{d\theta} \sum_{x=0}^{\infty} \theta(1 - \theta)^x &= \sum_{x=0}^{\infty} \frac{d}{d\theta} \theta(1 - \theta)^x \\ &= \sum_{x=0}^{\infty} [(1 - \theta)^x - \theta x(1 - \theta)^{x-1}] \\ &= \frac{1}{\theta} \sum_{x=0}^{\infty} \theta(1 - \theta)^x - \frac{1}{1 - \theta} \sum_{x=0}^{\infty} x\theta(1 - \theta)^x \end{aligned}$$

como $\sum_{x=0}^{\infty} \theta(1 - \theta)^x = 1$, para todo $0 < \theta < 1$, su derivada es cero,

$$\begin{aligned} 0 &= \frac{1}{\theta} - \frac{1}{1 - \theta} \sum_{x=0}^{\infty} x\theta(1 - \theta)^x \\ &= \frac{1}{\theta} - \frac{1}{1 - \theta} E[X] \end{aligned}$$

entonces $E[X] = \frac{1-\theta}{\theta}$.

Teorema 2.4.4 Suponga que la serie $\sum_{x=0}^{\infty} h(\theta, x)$ converge para todo θ en un intervalo (a, b) de números reales y:

- a. $\frac{\partial}{\partial \theta} h(\theta, x)$ es continua en θ para cada x ,
- b. $\sum_{x=0}^{\infty} \frac{\partial}{\partial \theta} h(\theta, x)$ converge uniformemente sobre cada subintervalo cerrado y acotado de (a, b) ,

entonces:

$$\frac{d}{d\theta} \sum_{x=0}^{\infty} h(\theta, x) = \sum_{x=0}^{\infty} \frac{\partial}{\partial \theta} h(\theta, x)$$

Ejemplo 2.4.4 (Continuación del ejemplo 2.4.3) Para aplicar el teorema 2.4.4 se identifica:

$$h(\theta, x) = \theta(1 - \theta)^x$$

y $\frac{\partial}{\partial \theta} h(\theta, x) = (1 - \theta)^x - \theta x(1 - \theta)^{x-1}$ para verificar que $\sum_{x=0}^{\infty} \frac{\partial}{\partial \theta} h(\theta, x)$ converge uniformemente. Se define $S_n(\theta)$ por:

$$S_n(\theta) = \sum_{x=0}^{\infty} \left[(1 - \theta)^x - \theta x(1 - \theta)^{x-1} \right]$$

La convergencia será uniforme en $[c, d] \subset (0, 1)$ si, dado $\epsilon > 0$, se puede encontrar N tal que:

$$n > N \Rightarrow |S_n(\theta) - S_{\infty}(\theta)| < \epsilon \text{ para todo } \theta \in [c, d]$$

Recordar la suma parcial de la serie geométrica. Si $y \neq 1$, entonces:

$$\sum_{k=0}^n y^k = \frac{1 - y^{n+1}}{1 - y}$$

luego:

$$\begin{aligned} \sum_{x=0}^n (1 - \theta)^x &= \frac{1 - (1 - \theta)^{n+1}}{\theta} \\ \sum_{x=0}^n \theta x(1 - \theta)^{x-1} &= \theta \sum_{x=0}^n -\frac{\partial}{\partial \theta} (1 - \theta)^x \\ &= -\theta \frac{d}{d\theta} \sum_{x=0}^n (1 - \theta)^x \\ &= -\theta \frac{d}{d\theta} \left[\frac{1 - (1 - \theta)^{n+1}}{\theta} \right] \\ &= \frac{(1 - (1 - \theta)^{n+1}) - (n + 1)\theta(1 - \theta)^n}{\theta} \end{aligned}$$

finalmente:

$$\begin{aligned} S_n(\theta) &= \frac{1 - (1 - \theta)^{n+1}}{\theta} - \frac{(1 - (1 - \theta)^{n+1}) - (n + 1)\theta(1 - \theta)^n}{\theta} \\ &= (n + 1)(1 - \theta)^n \end{aligned}$$

Es claro que, para $0 < \theta < 1$, $S_\infty = \lim_{n \rightarrow \infty} S_n(\theta) = 0$. Como $S_n(\theta)$ es continua la convergencia es uniforme sobre cualquier intervalo cerrado y acotado. Luego, las series de derivadas convergen uniformemente y el intercambio de la derivada con la integral está justificada.

Teorema 2.4.5 Suponga que la serie $\sum_{x=0}^{\infty} h(\theta, x)$ converge uniformemente en $[a, b]$ y que para cada x $h(\theta, x)$ es una función continua de θ , entonces:

$$\int_a^b \sum_{x=0}^{\infty} h(\theta, x) d\theta = \sum_{x=0}^{\infty} \int_a^b h(\theta, x) d\theta$$

Capítulo 3

Familias de distribuciones

3.1. Distribuciones discretas

Una variable aleatoria X se dice tiene una distribución discreta si su rango es numerable.

Distribución uniforme discreta

Una variable aleatoria X tiene *distribución uniforme discreta* $(1, N)$ si:

$$\Pr(X = x|N) = \frac{1}{N}, \quad x = 1, 2, \dots, N \quad (3.1.1)$$

donde N es un número entero. La media y varianza de la distribución uniforme discreta son:

$$E[X] = \frac{N+1}{2} \quad \text{Var}(X) = \frac{(N+1)(N-1)}{12}$$

Si el espacio muestral es cualquier rango de enteros $N_0, N_0 + 1, \dots, N_1$, entonces la función de probabilidad es:

$$\Pr(X = x|N_0, N_1) = \frac{1}{N_1 - N_0 + 1}$$

Distribución hipergeométrica

Una variable aleatoria X tiene *distribución hipergeométrica* (N, M, n) si:

$$\Pr(X = x|N, M, n) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \quad (3.1.2)$$

La media y varianza de la distribución hipergeométrica son:

$$E[X] = n \frac{M}{N} \quad \text{Var}(X) = \frac{nM}{N} \left(\frac{(N-M)(N-n)}{N(N-1)} \right)$$

Ejemplo 3.1.1 Un ingeniero de control de calidad inspecciona una muestra tomada al azar de dos calculadoras manuales de cada lote de tamaño 18. El lote será aceptado si ambas calculadoras están en buenas condiciones de trabajo de otra manera se inspecciona todo el lote y el costo se carga al vendedor. ¿Cuál es la probabilidad de que un lote se acepte sin tener que hacer una inspección total si éste contiene cuatro calculadoras que no están en buenas condiciones de trabajo? Aplicando la distribución hipergeométrica con $N = 18$, $M = 14$ y $n = 2$. Usando R, se tiene:

```
> dhyper(x = 2, m = 14, n = 4, k = 2)
[1] 0.5947712
```

Notar que R considera los parámetros de la distribución hipergeométrica (N, M, n) como $(m + n, m, k)$.

Distribución binomial

Una variable aleatoria tiene *distribución Bernoulli*(p) si:

$$X = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1 - p \end{cases}$$

con $0 \leq p \leq 1$. El valor de $X = 1$ es comúnmente llamado *éxito* y p la probabilidad de éxito. La media y varianza de la distribución Bernoulli son:

$$E[X] = p \quad \text{Var}(X) = p(1 - p)$$

Si se tiene una secuencia de variables aleatorias con distribución Bernoulli(p), es decir X_1, X_2, \dots, X_n tal que:

$$X_i = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1 - p \end{cases}$$

entonces la variable aleatoria $Y = \sum_{i=1}^n X_i$ tiene *distribución binomial*(n, p) y su función de probabilidad es :

$$\Pr(Y = y|n, p) = \binom{n}{y} p^y (1 - p)^{n-y}, \quad y = 0, 1, 2, \dots, n \quad (3.1.3)$$

La media y varianza de la distribución binomial son:

$$E[X] = np \quad \text{Var}(X) = np(1 - p)$$

Ejemplo 3.1.2 Suponga que se está interesado en calcular la probabilidad de obtener al menos un seis en cuatro lanzamientos de un dado regular. Este experimento puede ser modelado como una secuencia de cuatro ensayos de Bernoulli con probabilidad de éxito $p = 1/6$. Si se define la variable aleatoria como $X = \text{número de lanzamientos en los que se obtuvo seis}$, entonces $X \sim \text{binomial}(n = 4, p = 1/6)$. Usando R, se tiene:

```
> 1-dbinom(x = 0, size = 4, prob = 1/6)
[1] 0.5177469
```

Distribución Poisson

Una variable aleatoria X que toma valores en los enteros no negativos tiene *distribución Poisson*(λ) si:

$$\Pr(X = x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots \quad (3.1.4)$$

La media y varianza de la distribución Poisson son:

$$E[X] = \lambda \quad \text{Var}(X) = \lambda$$

Ejemplo 3.1.3 Considere un operador telefónico que recibe, en promedio, cinco llamadas cada tres minutos según un proceso de Poisson. ¿Cuál es la probabilidad de no recibir llamadas en el siguiente minuto? ¿Cuál es la

probabilidad de recibir al menos dos llamadas? Sea X = número de llamadas telefónicas recibidas en un minuto, entonces $X \sim \mathcal{P}(\lambda = 5/3)$. Usando R se tiene:

```
> dpois(x=0, lambda=5/3)
[1] 0.1888756
> 1-ppois(q=1, lambda=5/3)
[1] 0.4963317
```

Distribución binomial negativa

La distribución binomial cuenta el número de éxitos en un número fijo de ensayos de Bernoulli. Suponga que se desea contar el número de ensayos necesarios para obtener un número fijo de éxitos.

En una secuencia de ensayos independientes de Bernoulli(p) sea la variable aleatoria X definida como el ensayo en el que se obtiene el r -ésimo éxito, donde r es un entero, entonces:

$$\Pr(X = x|r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \dots \quad (3.1.5)$$

y se dice que X tiene *distribución binomial negativa*(r, p).

La distribución binomial negativa también se define en términos de la variable aleatoria Y = número de fracasos antes del r -ésimo éxito, es decir $Y = X - r$. Luego, la fórmula alternativa de la distribución binomial negativa es:

$$\Pr(Y = y|r, p) = \binom{r+y-1}{y} p^r (1-p)^y, \quad y = 0, 1, 2, \dots \quad (3.1.6)$$

A no ser que se indique lo contrario, nos referimos a la distribución binomial negativa(r, p) usando la función de probabilidad anterior. La media y varianza de la distribución binomial negativa son:

$$E[Y] = r \frac{(1-p)}{p} \quad \text{Var}(Y) = \frac{r(1-p)}{p^2}$$

Ejemplo 3.1.4 Una técnica conocida como muestreo binomial inverso suele usarse en poblaciones biológicas. Si la proporción de individuos que posee cierta característica es p y se realiza el proceso de muestreo hasta observar r de tales individuos, entonces el número de individuos muestreados es una variable aleatoria con distribución binomial negativa. Por ejemplo, suponga que en una población de moscas de la fruta se está interesado en la proporción que tienen alas vestigiales y se decide muestrear hasta encontrar 100 de estas moscas. La probabilidad que se tenga que examinar al menos N moscas es:

$$\begin{aligned}\Pr(X \geq N) &= \sum_{x=N}^{\infty} \binom{x-1}{99} p^{100} (1-p)^{x-100} \\ &= 1 - \sum_{x=100}^{N-1} \binom{x-1}{99} p^{100} (1-p)^{x-100}\end{aligned}$$

Dados p y N es posible evaluar la expresión anterior para determinar el valor más probable para la cantidad observada de moscas de la fruta.

Distribución geométrica

La distribución geométrica es un caso especial de la distribución binomial negativa. Si se toma $r = 1$ en 3.1.5 se tiene:

$$\Pr(X = x|p) = p(1-p)^{x-1}, \quad x = 1, 2, \dots \quad (3.1.7)$$

la que define la función de probabilidad de una variable aleatoria X con *distribución geométrica*(p) que se puede interpretar como el ensayo en el que se obtiene el primer éxito.

La media y varianza de X puede obtenerse usando las fórmulas de la binomial negativa cuando $X = Y + 1$. Luego:

$$\mathbb{E}[X] = \mathbb{E}[Y] + 1 = \frac{1}{p} \quad \text{Var}(X) = \frac{1-p}{p^2}$$

La distribución geométrica tiene una propiedad interesante conocida como la propiedad de *falta de memoria*. Para enteros $s > t$ se tiene:

$$\Pr(X > s|X > t) = \Pr(X > s - t) \quad (3.1.8)$$

Ejemplo 3.1.5 La distribución geométrica es usada para modelar el *tiempo de falla* de un componente. Por ejemplo, si la probabilidad que una bombilla

de luz falle en un día cualquiera es 0.01, entonces la probabilidad que la bombilla dure al menos 30 días es:

$$\Pr(X \geq 30) = \sum_{x=30}^{\infty} 0,001 (0,999)^{x-1}$$

y usando R, se tiene:

```
> 1-pgeom(q=29, prob=0.001)
[1] 0.970431
```

3.2. Distribuciones continuas

Distribución uniforme

La *distribución uniforme continua* se define sobre el intervalo $[a, b]$ con la siguiente función de densidad:

$$f(x|a, b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{de otro modo} \end{cases} \quad (3.2.1)$$

La media y varianza de la distribución uniforme continua son:

$$E[X] = \frac{a+b}{2} \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

Distribución gamma

La distribución gamma es una familia flexible de distribuciones sobre $[0, \infty)$. Si α es una constante positiva, la integral:

$$\int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

es finita. Solo si α es un entero positivo la integral anterior puede expresarse en forma cerrada. En cualquier caso su valor define la *función gamma*:

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (3.2.2)$$

Esta función satisface muchas relaciones útiles, en particular:

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha), \quad \alpha > 0 \quad (3.2.3)$$

Combinando 3.2.3 se tiene que para cualquier entero $n > 0$:

$$\Gamma(n) = (n - 1)! \quad (3.2.4)$$

como la integral en 3.2.2 es positiva, se tiene que:

$$f(t) = \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}, \quad 0 < t < \infty \quad (3.2.5)$$

es una función de densidad. Si se define la variable aleatoria $X = \beta T$, donde β es una constante positiva se obtiene la familia gamma(α, β):

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad 0 < x < \infty \quad (3.2.6)$$

donde $\alpha > 0$ es llamado el parámetro de forma y $\beta > 0$ el parámetro de escala. La media y varianza de la distribución gamma son:

$$E[X] = \alpha\beta \quad \text{Var}(X) = \alpha\beta^2$$

La función generatriz de momentos de la distribución gamma(α, β) es:

$$M_X(t) = \left(\frac{1}{1 - \beta t} \right)^\alpha$$

para $t < 1/\beta$. Existe una relación interesante entre la distribución gamma y Poisson. Si X es una variable aleatoria gamma(α, β), donde α es un entero, entonces para todo X ,

$$\Pr(X \leq x) = \Pr(Y \geq \alpha) \quad (3.2.7)$$

donde $Y \sim \mathcal{P}(x/\beta)$.

Existen dos casos importantes de la distribución gamma. Si se toma $\alpha = p/2$, donde p es un entero y $\beta = 2$, entonces la función de densidad gamma se convierte en:

$$f(x|p) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-x/2}, \quad 0 < x < \infty \quad (3.2.8)$$

que es conocida como la función de densidad *chi-cuadrado con p grados de libertad*.

Otro caso especial se obtiene cuando $\alpha = 1$, es decir:

$$f(x|\beta) = \frac{1}{\beta} e^{-x/\beta}, \quad 0 < x < \infty \quad (3.2.9)$$

conocida como la *función de densidad exponencial* con parámetro de escala β . La distribución exponencial comparte la propiedad de *pérdida de memoria* de la geométrica. Si $X \sim \mathcal{E}(\beta)$ entonces para $s > t \geq 0$:

$$\Pr(X > s|X > t) = \Pr(X > s - t)$$

Otra distribución relacionada a la familia gamma y exponencial es la *distribución Weibull*. Si $X \sim \mathcal{E}(\beta)$ entonces $Y = X^{1/\gamma}$ tiene distribución Weibull(γ, β):

$$f_Y(y|\gamma, \beta) = \frac{\gamma}{\beta} y^{\gamma-1} e^{-y^\gamma/\beta}, \quad 0 < y < \infty \quad (3.2.10)$$

donde $\gamma > 0$ y $\beta > 0$.

Distribución normal

La función de densidad de la distribución normal con media μ y varianza σ^2 es dada por:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \quad (3.2.11)$$

donde:

$$E[X] = \mu \quad \text{Var}(X) = \sigma^2$$

Si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces la variable aleatoria $Z = (X - \mu)/\sigma$ tiene distribución $\mathcal{N}(0, 1)$ conocida como *distribución normal estándar*.

Se puede probar que la función de densidad normal 3.2.11 toma su valor máximo en $x = \mu$ y que sus puntos de inflexión (donde la curva cambia de cóncava a convexo) son $x = \mu \pm \sigma$. La probabilidad que X este contenida en 1, 2 ó 3 desviaciones estándar alrededor de su media es:

$$\Pr(|X - \mu| \leq \sigma) = \Pr(|Z| \leq 1) = 0,6826$$

$$\Pr(|X - \mu| \leq 2\sigma) = \Pr(|Z| \leq 2) = 0,9544$$

$$\Pr(|X - \mu| \leq 3\sigma) = \Pr(|Z| \leq 3) = 0,9974$$

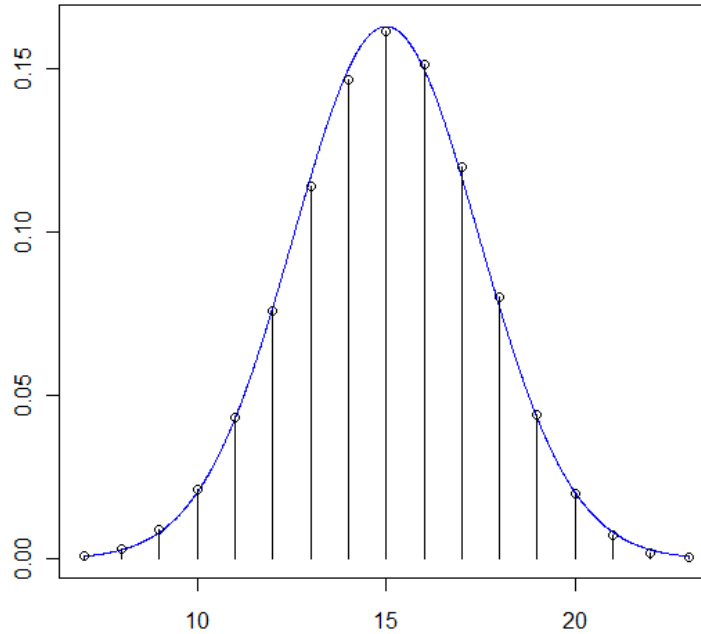
La distribución normal es usada como aproximación de otras distribuciones. Por ejemplo, si $X \sim \mathcal{BI}(n, p)$ entonces $E[X] = np$, $\text{Var}(X) = np(1 - p)$ y bajo ciertas condiciones favorables, la distribución de X puede aproximarse a la distribución normal con media $\mu = np$ y varianza $\sigma^2 = np(1 - p)$. Estas condiciones son n grande y p no debe estar cerca de 0 ó 1.

Si $X \sim \mathcal{BI}(n, p)$ y $Y \sim \mathcal{N}(np, np(1 - p))$ entonces:

$$\Pr(X \leq x) \approx \Pr(Y \leq x + 1/2)$$

$$\Pr(X \geq x) \approx \Pr(Y \geq x - 1/2)$$

Figura 3.1: Aproximación binomial a la distribución normal



Distribución beta

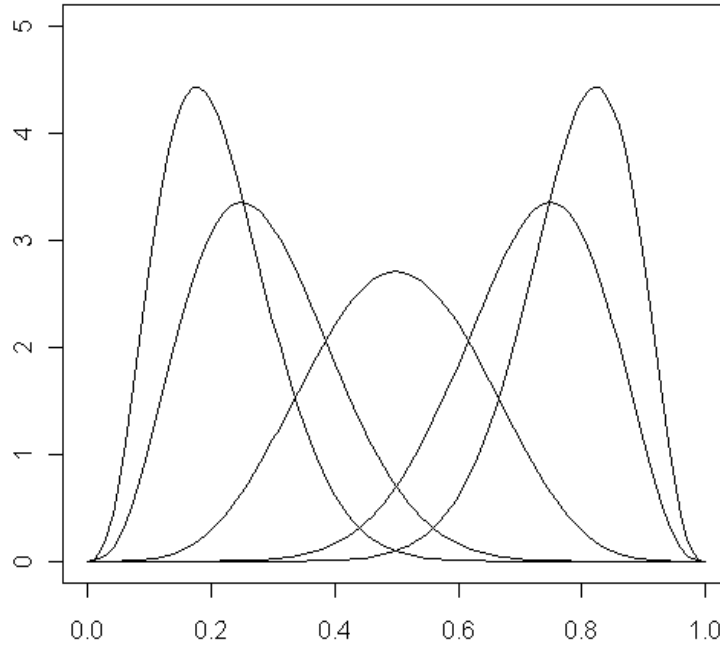
La familia de distribuciones beta es una familia continua en $(0, 1)$ indexada por dos parámetros. La *función de densidad* $\text{beta}(\alpha, \beta)$ es:

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1 \quad (3.2.12)$$

donde $\alpha > 0$, $\beta > 0$ y $B(\alpha, \beta)$ denota la *función beta*:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$$

Figura 3.2: Densidades beta



La función beta está relacionada con la función gamma a través de la siguiente identidad:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad (3.2.13)$$

El cálculo de los momentos para esta distribución se obtiene a través de la siguiente expresión:

$$E[X^n] = \frac{B(\alpha + n, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha + n)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + n)\Gamma(\alpha)} \quad (3.2.14)$$

Usando 3.2.3 y 3.2.14 con $n = 1$ y $n = 2$ se obtiene:

$$E[X] = \frac{\alpha}{\alpha + \beta} \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (3.2.15)$$

Distribución Cauchy

La *distribución Cauchy* es una distribución simétrica con forma de campana sobre $(-\infty, \infty)$ cuya función de densidad es:

$$f(x|\theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \quad -\infty < x < \infty \quad (3.2.16)$$

Tal como se mencionó en el capítulo 2, la media de la distribución de Cauchy no existe, esto es:

$$E[|X|] = \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{|x|}{1 + (x - \theta)^2} dx = \infty \quad (3.2.17)$$

por lo que tampoco existen los momentos para esta distribución. En particular, la función generatriz de momentos no existe.

Distribución log-normal

Si X es una variable aleatoria cuyo logaritmo tiene distribución normal, es decir $\log X \sim \mathcal{N}(\mu, \sigma^2)$ entonces X tiene *distribución log-normal*. La función de densidad de X se obtiene mediante una transformación directa de la función de densidad normal usando el teorema 2.1.2:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \quad (3.2.18)$$

$-\infty < \mu < \infty$ y $\sigma > 0$, conocida como la *función de densidad log-normal*. Los momentos de X pueden obtenerse usando 3.2.18 o la relación con la distribución normal:

$$E[X] = e^{\mu + \sigma^2/2} \quad \text{Var}(X) = e^{2(\mu + \sigma^2)} - e^{2\mu + \sigma^2} \quad (3.2.19)$$

Distribución doble exponencial

La *distribución doble exponencial* se toma reflejando la distribución exponencial alrededor de su media. Su función de densidad esta dada por:

$$f(x|\mu, \sigma) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}, \quad -\infty < x < \infty \quad (3.2.20)$$

$-\infty < \mu < \infty$ y $\sigma > 0$. La distribución doble exponencial proporciona una distribución simétrica con colas *más pesadas*. La media y varianza de esta distribución son:

$$E[X] = \mu \quad \text{Var}(X) = 2\sigma^2$$

3.3. Familias exponenciales

Una familia de funciones de densidad o probabilidad es llamada una *familia exponencial* si puede expresarse como:

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp \left\{ \sum_{i=1}^k w_i(\boldsymbol{\theta}) t_i(x) \right\} \quad (3.3.1)$$

donde $h(x) \geq 0$, $t_1(x), \dots, t_k(x)$ son funciones de las observaciones x , $c(\boldsymbol{\theta}) \geq 0$ y $w_1(\boldsymbol{\theta}), \dots, w_k(\boldsymbol{\theta})$ son funciones del vector de parámetros $\boldsymbol{\theta}$.

Muchas familias mencionadas en la sección anterior son familias exponenciales. Estas incluyen las familias continuas normal, gamma, beta y las familias discretas binomial, Poisson y binomial negativa.

Para verificar si una familia de funciones de probabilidad o densidad es una familia exponencial, se deben identificar las funciones $h(x)$, $c(\boldsymbol{\theta})$, $w_i(\boldsymbol{\theta})$ y $t_i(x)$ para demostrar que la familia tiene la forma 3.3.1.

Ejemplo 3.3.1 Considere la familia $\mathcal{BI}(n, p)$ con $0 < p < 1$ y n conocido.

Entonces la función de probabilidad es:

$$\begin{aligned}
 f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} \\
 &= \binom{n}{x} (1-p)^n \left(\frac{p}{1-p}\right)^x \\
 &= \binom{n}{x} (1-p)^n \exp \left\{ \log \left(\frac{p}{1-p} \right)^x \right\} \\
 &= \binom{n}{x} (1-p)^n \exp \left\{ \log \left(\frac{p}{1-p} \right) x \right\}
 \end{aligned}$$

Se define:

$$\begin{aligned}
 h(x) &= \binom{n}{x} \quad c(p) = (1-p)^n \\
 w_1(p) &= \log \left\{ \frac{p}{1-p} \right\} \quad y \quad t_1(x) = x
 \end{aligned}$$

entonces se tiene:

$$f(x|p) = h(x)c(p) \exp \{w_1(p)t_1(x)\} \quad (3.3.2)$$

que es de la forma 3.3.1 con $k = 1$.

Ejemplo 3.3.2 Sea $f(x|\mu, \sigma^2)$ la familia de distribuciones $\mathcal{N}(\mu, \sigma^2)$ donde $\theta = (\mu, \sigma^2)$, $-\infty < \mu < \infty$ y $\sigma > 0$. Entonces:

$$\begin{aligned}
 f(x|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{\mu^2}{2\sigma^2} \right\} \exp \left\{ -\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2} \right\}
 \end{aligned}$$

Se define:

$$h(x) = 1$$

$$\begin{aligned}
 c(\theta) &= c(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{\mu^2}{2\sigma^2} \right\} \\
 w_1(\mu, \sigma^2) &= \frac{1}{\sigma^2} \quad w_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2}
 \end{aligned}$$

$$t_1(x) = -x^2/2 \quad \text{y} \quad t_2(x) = x$$

luego $f(x|\mu, \sigma^2) = h(x)c(\mu, \sigma) \exp \{w_1(\mu, \sigma)t_1(x) + w_2(\mu, \sigma)t_2(x)\}$ tiene la forma 3.3.1 con $k = 2$.

El rango de la distribución no puede depender de θ en una familia exponencial. La definición completa de la función de probabilidad o densidad debe incluir el rango a través del uso de una función indicadora.

Definición 3.3.1 La *función indicadora* de un conjunto A , denotado por $I_A(x)$, es la función:

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Así en el ejemplo 3.3.2 la función de densidad normal puede ser escrita como:

$$f(x|\mu, \sigma^2) = h(x)c(\mu, \sigma) \exp \{w_1(\mu, \sigma)t_1(x) + w_2(\mu, \sigma)t_2(x)\} I_{(-\infty, \infty)}(x)$$

ya que la función indicadora depende solo de x , puede incorporarse en $h(x)$ mostrando que la función de densidad tiene la forma de una familia exponencial.

Ejemplo 3.3.3 La función de densidad dada por:

$$f(x|\theta) = \theta^{-1} \exp \{1 - (x/\theta)\}, \quad x > \theta \quad (3.3.3)$$

no es una familia exponencial aún cuando pueda escribirse como:

$$\exp \{1\} \theta^{-1} \exp \{\theta^{-1}x\} = h(x)c(\theta) \exp \{w(\theta)t(x)\}$$

donde $h(x) = \exp \{1\}$, $c(\theta) = \theta^{-1}$, $w(\theta) = \theta^{-1}$ y $t(x) = -x$, ya que:

$$f(x|\theta) = \theta^{-1} \exp \{1 - (x/\theta)\} I_{(\theta, \infty)}(x)$$

La función indicadora no puede incorporarse en ninguna de las funciones de 3.3.1 desde que no es función solamente de x o θ . Luego la función de densidad 3.3.3 no pertenece a una familia exponencial.

Una familia exponencial puede reparametrizarse como:

$$f(x|\eta) = h(x)c^*(\eta) \exp \left\{ \sum_{i=1}^k \eta_i t_i(x) \right\}$$

donde las funciones $h(x)$ y $t_i(x)$ son las mismas en la parametrización original
3.3.1. El conjunto:

$$\mathcal{H} = \eta = \left\{ (\eta_1, \dots, \eta_k) : \int_{-\infty}^{\infty} h(x) \exp \left\{ \sum_{i=1}^k \eta_i t_i(x) \right\} dx < \infty \right\}$$

es llamado el *espacio paramétrico natural* de la familia. La integral se reemplaza por la suma si X es discreta.

Para los valores de $\eta \in \mathcal{H}$, se tiene que:

$$c^*(\eta) = \left[\int_{-\infty}^{\infty} h(x) \exp \left\{ \sum_{i=1}^k \eta_i t_i(x) \right\} dx \right]^{-1}$$

lo cual asegura que la integral de la función de densidad sea uno. Luego, el conjunto:

$$\{\eta = (w_1(\theta), \dots, w_k(\theta)) : \theta \in \Theta\}$$

debe ser un subconjunto del espacio natural de parámetros.

Ejemplo 3.3.4 Para determinar el espacio natural paramétrico de la familia normal, se reemplaza $w_i(\mu, \sigma)$ con η_i en 3.3.2 y se obtiene:

$$\begin{aligned} f(x|\eta_1, \eta_2) &= \frac{\sqrt{\eta_1}}{\sqrt{2\pi}} \exp \left\{ -\frac{\eta_2^2}{2\eta_1} \right\} \exp \left\{ -\frac{\eta_1 x^2}{2} + \eta_2 x \right\} \\ &= \frac{\sqrt{\eta_1}}{\sqrt{2\pi}} \exp \left\{ -\frac{\eta_1}{2} \left(x - \frac{\eta_2}{\eta_1} \right)^2 \right\} \end{aligned}$$

La integral será finita si y solo si el término dentro de la función exponencial es negativo lo cual significa que η_1 debe ser positivo sin importar el valor que tome η_2 . Luego el espacio paramétrico natural es:

$$\{(\eta_1, \eta_2) : \eta_1 > 0, -\infty < \eta_2 < \infty\}$$

Identificando, se tiene que $\eta_2 = \mu/\sigma^2$ y $\eta_1 = 1/\sigma^2$.

3.4. Familias de locación y escala

En esta sección se discutirán tres técnicas para construir familias de distribuciones. Estos tipos son: familias de locación, escala y locación-escala. Cada familia se construye especificando una función de densidad simple, digamos $f(x)$, llamada la *función de densidad estándar* de la familia. Luego todas las funciones de densidad en la familia se generan transformando la función de densidad estándar en la forma prescrita.

Teorema 3.4.1 Sea $f(x)$ una función de densidad y sean μ y $\sigma > 0$ constantes, entonces:

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$

es una función de densidad.

Definición 3.4.1 Sea $f(x)$ una función de densidad. Entonces la familia de funciones de densidad $f(x - \mu)$ indexada por el parámetro $-\infty < \mu < \infty$, $-\infty < \mu < \infty$, es llamada la *familia de locación con función de densidad estándar* $f(x)$ y μ es el *parámetro de locación* para la familia.

Ejemplo 3.4.1 Sea $f(x) = e^{-x}$, $x \geq 0$. Para formar la familia de locación se reemplaza x con $x - \mu$:

$$f(x|\mu) = \begin{cases} e^{-(x-\mu)} & x - \mu \geq 0 \\ 0 & x - \mu < 0 \end{cases} = \begin{cases} e^{-(x-\mu)} & x \geq \mu \\ 0 & x < \mu \end{cases}$$

Definición 3.4.2 Sea $f(x)$ una función de densidad. Entonces para todo $\sigma > 0$, la familia de funciones de densidad $(1/\sigma)f(x/\sigma)$ indexada por el parámetro σ , es llamada la *familia de escala* con función de densidad estándar $f(x)$ y σ es el *parámetro de escala* de la familia.

Ejemplo 3.4.2 La distribución gamma con α fijo y la distribución normal con $\mu = 0$ son ejemplos de familias de escala. En cada caso la función de densidad estándar se obtiene tomando el parámetro de escala igual a 1.

Definición 3.4.3 Sea $f(x)$ una función de densidad. Entonces la familia de funciones de densidad $(1/\sigma)f((x - \mu)/\sigma)$, $-\infty < \mu < \infty$ y $\sigma > 0$, indexada por los parámetros (μ, σ) es llamada la *familia de locación-escala* con función

de densidad estándar $f(x)$, μ es el *parámetro de locación* y σ es el *parámetro de escala*.

Ejemplo 3.4.3 La distribución normal y la distribución exponencial son ejemplos de familias de locación-escala.

Teorema 3.4.2 Sea $f(\cdot)$ una función de densidad. Si $\mu \in \Re$ y $\sigma > 0$, entonces X es una variable aleatoria con función de densidad $(1/\sigma)f((x - \mu)/\sigma)$ si y solo si existe una variable aleatoria Z con función de densidad $f(z)$ y $X = \sigma Z + \mu$.

Prueba: Se define $z = g(x) = (x - \mu)/\sigma$ entonces $g^{-1}(z) = \sigma z + \mu$ y $|dg^{-1}(z)/dz| = \sigma$. Aplicando el teorema 2.1.1 la función de densidad de Z es:

$$f_Z(z) = f_X(g^{-1}(z)) \left| \frac{d}{dz} g^{-1}(z) \right| = \frac{1}{\sigma} f\left(\frac{(\sigma z + \mu) - \mu}{\sigma}\right) \sigma = f(z)$$

Se define $x = g(z) = \sigma z + \mu$ entonces $g^{-1}(x) = (x - \mu)/\sigma$, y $|dg^{-1}(x)/dx| = 1/\sigma$. Aplicando el teorema 2.1.1 la función de densidad de X es:

$$f_X(x) = f_Z(g^{-1}(x)) \left| \frac{d}{dx} g^{-1}(x) \right| = f\left(\frac{x - \mu}{\sigma}\right) \frac{1}{\sigma}$$

Teorema 3.4.3 Sea Z una variable aleatoria con función de densidad $f(z)$. Suponga que $E[Z]$ y $\text{Var}(Z)$ existen. Si X es una variable aleatoria con función de densidad $(1/\sigma)f((x - \mu)/\sigma)$, entonces:

$$E[X] = \sigma E[Z] + \mu \text{ y } \text{Var}(X) = \sigma^2 \text{Var}(Z)$$

En particular si, $E[Z] = 0$ y $\text{Var}(Z) = 1$ entonces $E[X] = \mu$ y $\text{Var}(X) = \sigma^2$.

El cálculo de una probabilidad para cualquier miembro de una familia de locación escala puede obtenerse en términos de la variable estandarizada Z ya que:

$$\Pr(X \leq x) = \Pr\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Pr\left(Z \leq \frac{x - \mu}{\sigma}\right)$$

El cálculo de las probabilidades para la distribución normal usando la distribución normal estándar es un claro ejemplo.

3.5. Desigualdades e identidades

3.5.1. Desigualdades para probabilidades

Teorema 3.5.1 (Desigualdad de Chebyshev) Sea X una variable aleatoria y sea $g(x)$ una función no negativa. Entonces, para todo $r > 0$:

$$\Pr(g(X) \geq r) \leq \frac{\mathbb{E}[g(X)]}{r}$$

Ejemplo 3.5.1 Si Z tiene distribución normal estándar, entonces para todo $t > 0$:

$$\Pr(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}$$

3.5.2. Identidades

Teorema 3.5.2 Sea $X_{\alpha,\beta}$ denota una variable aleatoria con distribución $\mathcal{G}(\alpha, \beta)$ con función de densidad $f(x/\alpha, \beta)$, donde $\alpha > 1$. Entonces para constantes cualesquiera a y b :

$$\Pr(a < X_{\alpha,\beta} < b) = \beta (f(a/\alpha, \beta) - f(b/\alpha, \beta)) + \Pr(a < X_{\alpha-1,\beta} < b)$$

Lema 3.5.1 (Lema de Stein) Sea $X \sim \mathcal{N}(\theta, \sigma^2)$ y sea g una función diferenciable que satisface $\mathbb{E}[g'(X)] < \infty$, entonces:

$$\mathbb{E}[g(X)(X - \theta)] = \sigma^2 \mathbb{E}[g'(X)]$$

Teorema 3.5.3 Sea χ_p^2 que denota una variable aleatoria con distribución chi-cuadrado con p grados de libertad. Para toda función $h(x)$, siempre que su valor esperado exista:

$$\mathbb{E}[h(\chi_p^2)] = p \mathbb{E}\left[\frac{h(\chi_{p+2}^2)}{\chi_{p+2}^2}\right]$$

Capítulo 4

Variables aleatorias múltiples

4.1. Distribución conjunta y marginal

Definición 4.1.1 Un *vector aleatorio n -dimensional* es una función que va de un espacio muestral S a un espacio euclideo n -dimensional \mathbb{R}^n .

Ejemplo 4.1.1 Considere el experimento de lanzar 2 veces un dado. El espacio muestral para este experimento tiene 36 puntos igualmente probables. Sean las variables aleatorias X = suma de los dados, y Y = |diferencia de los dados|, entonces el vector aleatorio (X, Y) es llamado *vector aleatorio discreto* ya que solo tiene un número finito de posibles valores.

Definición 4.1.2 Sea (X, Y) un vector aleatorio discreto bivariado. Entonces la función $f(X, Y)$, que va de \mathbb{R}^2 hacia \mathbb{R} , definido por $f(x, y) = \Pr(X = x, Y = y)$ es llamado la *función de probabilidad conjunta* de (X, Y) .

Tabla 4.1: Valores de la función de probabilidad conjunta $f(x, y)$

		x										
		2	3	4	5	6	7	8	9	10	11	12
y	0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$
	1		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$	
	2			$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		
	3				$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$			
	4					$\frac{1}{18}$		$\frac{1}{18}$				
	5						$\frac{1}{18}$					

La función de probabilidad conjunta puede usarse en el cálculo de probabilidades para cualquier evento definido en términos de (X, Y) . Sea A un subconjunto de \mathbb{R}^2 . Entonces:

$$\Pr((X, Y) \in A) = \sum_{(X, Y) \in A} f(x, y)$$

Ejemplo 4.1.2 Para (X, Y) cuya función de probabilidad conjunta se encuentra en la tabla 4.1, suponga que $A = \{(x, y) : x = 7, y \leq 4\}$, entonces:

$$\Pr(X = 7, Y \leq 4) = \Pr((X, Y) \in A) = f(7, 1) + f(7, 3) = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}$$

Sea $g(x, y)$ una función de valor real definido para todos los posibles valores (x, y) de vector aleatorio discreto (X, Y) . Entonces $g(X, Y)$ es también una variable aleatoria y su valor esperado se define por:

$$E[g(X, Y)] = \sum_{(x, y) \in \mathbb{R}^2} g(x, y) f(x, y)$$

Ejemplo 4.1.3 Usando la tabla 4.1 el valor esperado de $g(X, Y) = XY$,

$$E[g(X, Y)] = \sum xyf(x, y) = (2)(0)\frac{1}{36} + \cdots + (7)(5)\frac{1}{18} = 13\frac{11}{18}$$

Las propiedades vistas en el teorema 2.2.1 son válidas al reemplazar x por (x, y) . Por ejemplo si $g_1(x, y)$, $g_2(x, y)$ son dos funciones; a , b y c son constantes, entonces:

$$E[ag_1(X, Y) + bg_2(X, Y) + c] = aE[g_1(X, Y)] + bE[g_2(X, Y)] + c$$

La función de probabilidad conjunta del vector aleatorio (X, Y) debe cumplir con $f(x, y) \geq 0$, para todo (x, y) . Además:

$$\sum_{(x, y) \in \mathbb{R}^2} f(x, y) = \Pr((X, Y) \in \mathbb{R}^2) = 1.$$

Ejemplo 4.1.4 Se define $f(w, z)$ por:

$$f(0, 0) = f(0, 1) = \frac{1}{6}$$

$$f(1, 0) = f(1, 1) = \frac{1}{3}$$

$$f(w, z) = 0 \text{ para cualquier otro } (w, z)$$

Si S es el espacio muestral formado por los 36 puntos que resultan de lanzar dos dados. Sean:

$$W = \begin{cases} 0 & \text{si el 1ero es como máximo 2} \\ 1 & \text{si el 1ero es más de 2} \end{cases}$$

$$Z = \begin{cases} 0 & \text{si el 2do es impar} \\ 1 & \text{si el 2do es par} \end{cases}$$

Luego $f(w, z)$ es la función de probabilidad conjunta para (W, Z) .

Teorema 4.1.1 Sea (X, Y) un vector aleatorio discreto cuya función de probabilidad conjunta es $f(x, y)$, entonces la función de probabilidad marginal de X , $f_X(x) = \Pr(X = x)$, y de Y , $f_Y(y) = \Pr(Y = y)$, están dadas por:

$$f_X(x) = \sum_{y \in \mathfrak{R}} f(x, y) \quad \text{y} \quad f_Y(y) = \sum_{x \in \mathfrak{R}} f(x, y)$$

Prueba: Para todo $x \in \mathfrak{R}$, sea $A_x = \{(x, y) : -\infty < y < \infty\}$. Es decir, A_x es la línea en el plano con primer coordenada igual a x . Entonces:

$$\begin{aligned} f_X(x) &= \Pr(X = x) \\ &= \Pr(X = x, -\infty < y < \infty) \\ &= \Pr((X, Y) \in A_x) \\ &= \sum_{(X, Y) \in A_x} f(x, y) \\ &= \sum_{y \in \mathfrak{R}} f(x, y) \end{aligned}$$

Ejemplo 4.1.5 Usando el teorema 4.1.1 se puede calcular las distribuciones marginales de X e Y a partir de la distribución conjunta de la tabla 4.1:

$$f_Y(0) = \frac{1}{6} \quad f_Y(1) = \frac{5}{18} \quad f_Y(2) = \frac{2}{9} \quad f_Y(3) = \frac{1}{6} \quad f_Y(4) = \frac{1}{9} \quad f_Y(5) = \frac{1}{18}$$

Ejemplo 4.1.6 Usando la función de probabilidad marginal de Y se puede calcular:

$$\Pr(Y < 3) = \frac{2}{3} \quad E[Y^3] = 20\frac{11}{18}$$

Definición 4.1.3 Una función $f(x, y)$ que va de \mathbb{R}^2 hacia \mathbb{R} es llamada *función de densidad conjunta del vector aleatorio bivariado continuo* (X, Y) si, para todo $A \subset \mathbb{R}^2$:

$$\Pr((X, Y) \in A) = \iint_A f(x, y) dx dy$$

Si $g(x, y)$ es una función de valor real, entonces *el valor esperado* de $g(X, Y)$ se define por:

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy \quad (4.1.1)$$

Las *funciones de densidad marginales* de X y Y son definidas, reemplazando las sumatorias por las integrales. Estas funciones pueden usarse para calcular probabilidades o valores esperados que involucran solo a X o Y . Simplificando, las funciones de densidad marginales de X y Y son definidas por:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy, \quad -\infty < x < \infty \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx, \quad -\infty < y < \infty \end{aligned} \quad (4.1.2)$$

Toda función $f(x, y)$ que satisface $f(x, y) \geq 0$, para todo $(X, Y) \in \mathbb{R}^2$, y:

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

se dice que es la función de densidad conjunta para algún vector aleatorio bivariado (X, Y) .

Ejemplo 4.1.7 Se define la función de densidad conjunta por:

$$f(x, y) = \begin{cases} 6xy^2 & 0 < x < 1, \ 0 < y < 1 \\ 0 & \text{de otro modo} \end{cases}$$

Si se desea calcular $\Pr(X + Y \geq 1)$, sea $A = \{(x, y) : x + y \geq 1\}$ entonces:

$$\Pr(X + Y \geq 1) = \iint_A f(x, y) dx dy = \int_{y=0}^{y=1} \int_{x=1-y}^{x=1} 6xy^2 dx dy = \frac{9}{10}$$

Usando 4.1.2 puede obtenerse la función de densidad marginal de X y Y :

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dx dy = \int_{y=0}^{y=1} 6xy^2 dy = 2xy^3 \Big|_{y=0}^{y=1} = 2x$$

Esta función de densidad de X puede ser usada para calcular probabilidades, por ejemplo:

$$\Pr(1/2 < X < 3/4) = \int_{x=1/2}^{x=3/4} 2x dx = \frac{5}{10}$$

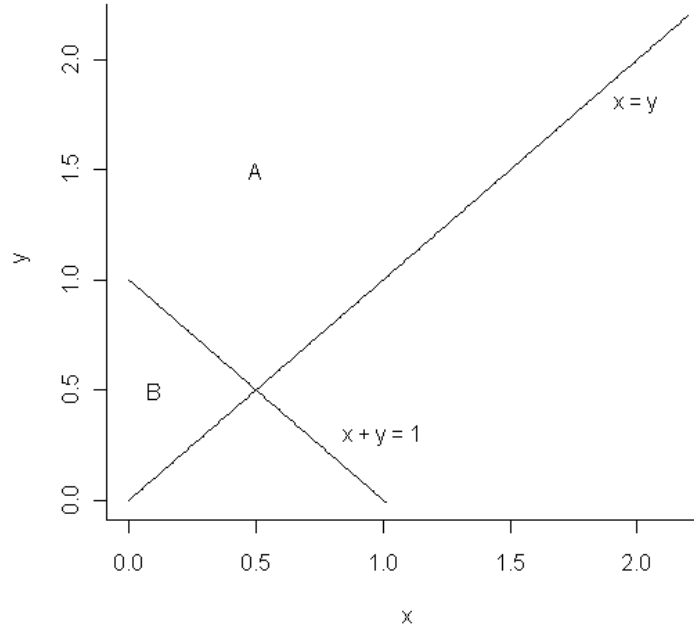
Ejemplo 4.1.8 Sea $f(x, y) = e^{-y}, 0 < x < y$. Aparentemente $f(x, y)$ no depende de X sin embargo:

$$f(x, y) = e^{-y} I_{\{0 < x < y\}}(x, y)$$

Para calcular $\Pr(X + Y \geq 1)$ se define $A = \{(x, y) : x + y \geq 1\}$. El proceso de integración es mucho más sencillo si se hace sobre el conjunto $B = \{(x, y) : x + y < 1\}$:

$$\begin{aligned} \Pr(X + Y \geq 1) &= 1 - \Pr(X + Y < 1) \\ &= 1 - \int_{x=0}^{x=1/2} \int_{y=x}^{y=1-x} e^{-y} dy dx \\ &= 1 - \int_0^{1/2} (e^{-x} - e^{-(1-x)}) dx \\ &= 0,845182 \end{aligned}$$

Figura 4.1: Región para el ejemplo 4.1.8



La función de densidad conjunta de (X, Y) puede describirse completamente usando la función de distribución acumulada conjunta $F(x, y)$ definida por:

$$F(x, y) = \Pr(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds$$

para todo $(x, y) \in \mathbb{R}^2$. Usando el teorema fundamental del cálculo:

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y) \quad (4.1.3)$$

para todos los puntos de continuidad de $f(x, y)$.

4.2. Distribuciones condicionales e independencia

Definición 4.2.1 Sea (X, Y) un vector aleatorio bivariado discreto con función de probabilidad conjunta $f(x, y)$ y funciones de probabilidad marginales

$f_X(x)$ y $f_Y(y)$. Para todo x tal que $f_X(x) > 0$, la *función de probabilidad condicional de Y dado $X = x$* es la función de y denotada por $f(y|x)$ y definida por:

$$f(y|x) = \Pr(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}$$

Para todo y tal que $f_Y(y) > 0$, la *función de probabilidad condicional de x dado que $Y = y$* es la función de x denotada por $f(x|y)$ y definida por:

$$f(x|y) = \Pr(X = x|Y = y) = \frac{f(x, y)}{f_Y(y)}$$

Ejemplo 4.2.1 Se define la función de probabilidad conjunta de (X, Y) por:

	y			
x	10	20	30	$f_X(x)$
0	$\frac{2}{18}$	$\frac{2}{18}$	0	$\frac{4}{18}$
1	$\frac{3}{18}$	$\frac{4}{18}$	$\frac{3}{18}$	$\frac{10}{18}$
2	0	0	$\frac{4}{18}$	$\frac{4}{18}$
$f_Y(y)$	$\frac{5}{18}$	$\frac{6}{18}$	$\frac{7}{18}$	

$$f(Y = y|X = 0) = \begin{cases} \frac{1}{2} & \text{si } y = 10 \\ \frac{1}{2} & \text{si } y = 20 \\ 0 & \text{de otro modo} \end{cases}$$

$$f(Y = y|X = 1) = \begin{cases} \frac{3}{10} & \text{si } y = 10 \\ \frac{4}{10} & \text{si } y = 20 \\ \frac{3}{10} & \text{si } y = 30 \\ 0 & \text{de otro modo} \end{cases}$$

$$f(Y = y|X = 2) = \begin{cases} 1 & \text{si } y = 30 \\ 0 & \text{de otro modo} \end{cases}$$

También puede calcularse probabilidades condicionales, por ejemplo:

$$\Pr(Y > 10|X = 1) = f(20|1) + f(30|1) = \frac{7}{10}$$

$$\Pr(Y > 10|X = 0) = f(20|0) = \frac{1}{2}$$

Definición 4.2.2 Sea (X, Y) un vector aleatorio continuo bivariado con función de densidad conjunta $f(x, y)$ y funciones de densidad marginales $f_X(x)$ y $f_Y(y)$. Para todo x tal que $f_X(x) > 0$, la *función de densidad condicional de Y dado que $X = x$* es la función de y denotada por $f(y|x)$ y definida por:

$$f(y|x) = \frac{f(x, y)}{f_X(x)}$$

Para todo y tal que $f_Y(y) > 0$, la *función de densidad condicional de X dado que $Y = y$* es la función de x denotada por $f(x|y)$ y definida por:

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

Si $g(Y)$ es una función de Y , entonces *el valor esperado condicional de $g(Y)$ dado que $X = x$* se denota por $E[g(Y)|x]$ y se define por:

$$E[g(Y)|x] = \sum_y g(y)f(y|x) \quad \text{y} \quad E[g(Y)|x] = \int_{-\infty}^{\infty} g(y)f(y|x)dy$$

El valor esperado condicional tiene todas las propiedades del valor esperado vistas en el teorema 2.2.1.

Ejemplo 4.2.2 Como en el ejemplo 4.1.8, sea un vector aleatorio continuo (X, Y) con función de densidad conjunta $f(x, y) = e^{-y}$, $0 < x < y$. Suponga se desea calcular la función de densidad condicional de Y dado $X = x$. La función de densidad marginal de X se puede calcular como sigue:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy = \int_{y=x}^{y=\infty} e^{-y}dy = e^{-x}, \quad x > 0$$

entonces $X \sim \mathcal{E}(\beta = 1)$. Luego:

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{-(y-x)}, \quad y > x$$

Dado $X = x$, la distribución de Y es exponencial donde x es el parámetro de locación y $\beta = 1$ es el parámetro de escala. La distribución condicional de Y es diferente para cada valor de x . Además:

$$E[Y|X = x] = \int_{y=x}^{y=\infty} ye^{-(y-x)} dy = 1 + x$$

La variancia de la función de densidad $f(y|x)$ es llamada *variancia condicional de Y dado X = x*. Usando la notación $\text{Var}(Y|X)$ se tiene:

$$\begin{aligned} \text{Var}(Y|X) &= E[Y^2|X] - (E[Y|X])^2 \\ &= \int_x^\infty y^2 e^{-(y-x)} dy - \left(\int_x^\infty ye^{-(y-x)} dy \right)^2 \\ &= 1 \end{aligned}$$

En este caso la variancia condicional de Y dado $X = x$ es la misma para todos los valores de x . Esta variancia condicional puede compararse con la *variancia no condicional de Y*. La distribución marginal de Y es $\mathcal{G}(2, 1)$, la cual tiene $\text{Var}(Y) = 2$. Dado el valor $X = x$, la variabilidad en Y se reduce considerablemente.

Definición 4.2.3 Sea (X, Y) un vector aleatorio bivariado con función de probabilidad o función de densidad conjunta $f(x, y)$ y funciones de probabilidad o densidad marginales $f_X(x)$ y $f_Y(y)$. Entonces X y Y son llamadas *variables aleatorias independientes* si, para todo $x \in \mathfrak{R}$ y $y \in \mathfrak{R}$,

$$f(x, y) = f_X(x)f_Y(y) \quad (4.2.1)$$

Si X y Y son independientes, la función de probabilidad o densidad condicional Y dado $X = x$ es:

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x)$$

para cualquier valor de x . Así, para todo $A \subset \mathfrak{R}$ y $x \in \mathfrak{R}$,

$$\Pr(Y \in A|x) = \int_A f(y|x) dy = \int_A f_Y(y) dy = \Pr(Y \in A)$$

El saber que $X = x$ no brinda información adicional acerca de Y .

Ejemplo 4.2.3 Sea el vector aleatorio bivariado discreto (X, Y) con función de probabilidad conjunta dada por:

x	y			$f_X(x)$
	1	2	3	
10	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{5}{10}$
20	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{5}{10}$
$f_Y(y)$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$	

Las variables aleatorias X y Y no son independientes debido a que 4.2.1 no se cumple para todo (x, y) . Por ejemplo:

$$f(3, 10) = \frac{1}{5} \neq \frac{1}{2} \times \frac{1}{2} = f_X(3)f_Y(10)$$

Notar que $f(10, 1) = \frac{1}{10} = \frac{1}{2} \times \frac{1}{5} = f_X(10)f_Y(1)$. El hecho de que 4.2.1 sea verdadera para algunos valores de (x, y) no aseguran que X y Y sean independientes.

Lema 4.2.1 Sea (X, Y) un vector aleatorio bivariado con función de probabilidad o densidad conjunta $f(x, y)$. Entonces X y Y son variables aleatorias independientes sí y solo si existen funciones $g(x)$ y $h(y)$ tales que, para todo $x \in \mathfrak{R}$ y $y \in \mathfrak{R}$,

$$f(x, y) = g(x)h(y)$$

Prueba: Si se define $g(x) = f_X(x)$ y $h(y) = f_Y(y)$ y usando 4.2.1 es fácil probar una de las direcciones. Para probar la otra dirección, suponga que $f(x, y) = g(x)h(y)$. Se define $\int_{-\infty}^{\infty} g(x)dx = c$ y $\int_{-\infty}^{\infty} h(y)dy = d$, donde las constantes c y d satisfacen:

$$cd = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dx dy \quad (4.2.2)$$

Además, las funciones de densidad marginales están dadas por:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} g(x)h(y)dy = g(x)d \\ f_Y(y) &= \int_{-\infty}^{\infty} g(x)h(y)dx = h(y)c \end{aligned} \quad (4.2.3)$$

Luego, usando 4.2.2 y 4.2.3, se tiene:

$$f(x, y) = g(x)h(y) = g(x)h(y)cd = f_X(x)f_Y(y)$$

demostrando que X y Y son independientes. Reemplazando las integrales por sumatorias se prueba el lema para vectores aleatorios discretos.

Ejemplo 4.2.4 Considere la función de densidad conjunta:

$$f(x, y) = \frac{1}{384} x^2 y^4 e^{-y-(x/2)}, \quad x > 0, y > 0$$

Si se definen:

$$g(x) = \begin{cases} x^2 e^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$h(y) = \begin{cases} y^4 e^{-y}/384 & y > 0 \\ 0 & y \leq 0 \end{cases}$$

entonces $f(x, y) = g(x)h(y)$ para todo $x \in \Re$ y $y \in \Re$. Por el lema 4.2.1, se concluye que X y Y son variables aleatorias independientes. Notar que no fué necesario calcular las funciones de densidad marginales.

Teorema 4.2.1 Sean X y Y variables aleatorias independientes:

- a. Para todo $A \subset \Re$ y $B \subset \Re$, $\Pr(X \in A, Y \in B) = \Pr(X \in A) \Pr(Y \in B)$, esto es, los eventos $\{X \in A\}$ y $\{Y \in B\}$ son independientes.
- b. Sea $g(x)$ una función que depende sólo de x y $h(y)$ una función que depende sólo y . Entonces:

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Prueba: Notar que:

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f(x, y)dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dx dy \\ &= \int_{-\infty}^{\infty} h(y)f_Y(y) \int_{-\infty}^{\infty} g(x)f_X(x)dx dy \\ &= \left(\int_{-\infty}^{\infty} g(x)f_X(x)dx \right) \left(\int_{-\infty}^{\infty} h(y)f_Y(y)dy \right) \\ &= E[g(X)]E[h(Y)] \end{aligned}$$

Sea $g(x)$ la función indicadora del conjunto A y sea $h(y)$ la función indicadora del conjunto B . Notar que $g(x)h(y)$ es la función indicadora del conjunto $C \subset \mathbb{R}^2$ definido por $C = \{(x, y) : x \in A, y \in B\}$. Notar que para una función indicadora como $g(x)$, $E[g(X)] = \Pr(X \in A)$. Usando el resultado anterior se tiene:

$$\begin{aligned}\Pr(X \in A, Y \in B) &= \Pr((X, Y) \in C) = E[g(X)h(Y)] \\ &= E[g(X)]E[h(Y)] = \Pr(X \in A)\Pr(Y \in B)\end{aligned}$$

Ejemplo 4.2.5 Sean X y Y variables aleatorias independientes con distribución $\mathcal{E}(1)$. Por el teorema 4.3.2 se tiene:

$$\Pr(X \geq 4, Y < 3) = \Pr(X \geq 4)\Pr(Y < 3)$$

Sean $g(x) = x^2$ y $h(y) = y$, se tiene que:

$$E[X^2Y] = E[X^2]E[Y] = (\text{Var}(X) + E[X]^2)E[Y] = (1 + 1^2)1 = 2$$

Teorema 4.2.2 Sean X y Y variables aleatorias independientes con funciones generatrices de momentos $M_X(t)$ y $M_Y(t)$ respectivamente. Entonces la función generatriz de momentos de la variable aleatoria $Z = X + Y$ es:

$$M_Z(t) = M_X(t)M_Y(t)$$

Prueba: Usando la definición de función generatriz de momentos:

$$M_Z(t) = E[e^{tZ}] = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$$

Ejemplo 4.2.6 Algunas veces el teorema 4.2.2 puede usarse para encontrar fácilmente la distribución de $Z = X + Y$ a partir de las distribuciones de X y Y . Por ejemplo, sea $X \sim \mathcal{N}(\mu, \sigma^2)$ y $Y \sim \mathcal{N}(\gamma, \tau^2)$ variables aleatorias independientes. Las funciones generatrices de momentos de X y Y son:

$$M_X(t) = \exp\left\{\mu t + \sigma^2 t^2/2\right\} \quad \text{y} \quad M_Y(t) = \exp\left\{\gamma t + \tau^2 t^2/2\right\}$$

Luego:

$$M_Z(t) = M_X(t)M_Y(t) = \exp\left\{(\mu + \gamma)t + (\sigma^2 + \tau^2)t^2/2\right\}$$

La función generatriz de momentos de Z corresponde a una variable aleatoria con distribución normal con media $(\mu + \gamma)$ y varianza $(\sigma^2 + \tau^2)$.

4.3. Transformaciones bivariadas

Sea (X, Y) un vector aleatorio bivariado con una distribución conocida. Considere un nuevo vector aleatorio bivariado (U, V) definido por $U = g_1(X, Y)$ y $V = g_2(X, Y)$ donde $g_1(x, y)$ y $g_2(x, y)$ son funciones específicas. Si \mathcal{B} es cualquier subconjunto de \mathbb{R}^2 , entonces $(U, V) \in \mathcal{B}$ sí y solo si $(X, Y) \in \mathcal{A}$, donde $\mathcal{A} = \{(x, y) : (g_1(x, y), g_2(x, y)) \in \mathcal{B}\}$. Luego $\Pr((U, V) \in \mathcal{B}) = \Pr(X, Y) \in \mathcal{A}$ y la distribución conjunta de (U, V) se determina completamente usando la distribución conjunta de (X, Y) .

4.3.1. Caso discreto

Si (X, Y) es un vector aleatorio bivariado discreto, entonces existe solo un conjunto numerable de valores para los que la función de probabilidad conjunta de (X, Y) es positiva, digamos el conjunto \mathcal{A} . Se define el conjunto $\mathcal{B} = \{(u, v) : u = g_1(x, y), v = g_2(x, y) \text{ para algún } (x, y) \in \mathcal{A}\}$. Entonces \mathcal{B} es el conjunto numerable de posibles valores para el vector aleatorio discreto (U, V) . Si para todo $(u, v) \in \mathcal{B}$, \mathcal{A}_{uv} se define como $\{(x, y) \in \mathcal{A} : g_1(x, y) = u, g_2(x, y) = v\}$ entonces la función de probabilidad conjunta de (U, V) , $f_{U,V}(u, v)$, puede calcularse a partir de la función de probabilidad conjunta de (X, Y) por:

$$f_{U,V}(u, v) = \Pr((x, y) \in \mathcal{A}_{uv}) = \sum_{(x,y) \in \mathcal{A}_{uv}} f_{X,Y}(x, y) \quad (4.3.1)$$

Ejemplo 4.3.1 Sean $X \sim \mathcal{P}(\theta)$ y $Y \sim \mathcal{P}(\lambda)$ variables aleatorias independientes. La función de probabilidad conjunta de (X, Y) es:

$$f_{X,Y}(x, y) = \frac{\theta^x e^{-\theta}}{x!} \frac{\lambda^y e^{-\lambda}}{y!}, \quad x = 0, 1, 2, \dots; y = 0, 1, 2, \dots$$

El conjunto $\mathcal{A} = \{(x, y) : x = 0, 1, 2, \dots; y = 0, 1, 2, \dots\}$. Se define $U = X + Y$ y $V = Y$. El conjunto $\mathcal{B} = \{(u, v) : v = 0, 1, \dots; u = v, v + 1, \dots\}$. Para todo $(u, v) \in \mathcal{B}$, el único valor (x, y) que satisface $x + y = u$ y $y = v$ es $x = u - v$ e $y = v$. El conjunto \mathcal{A}_{uv} consiste de los puntos $(u - v, v)$, entonces:

$$f_{U,V}(u, v) = f_{X,Y}(u - v, v) = \frac{\theta^{u-v} e^{-\theta}}{(u-v)!} \frac{\lambda^v e^{-\lambda}}{v!}, \quad v = 0, 1, \dots; u = v, v + 1, \dots$$

En este ejemplo sería interesante calcular la función de probabilidad marginal de U . Para todo entero no negativo u , $f_{U,V}(u, v) > 0$ solo si $v = 0, 1, \dots, u$. Es decir:

$$f_U(u) = \sum_{v=0}^u \frac{\theta^{u-v} e^\theta}{(u-v)!} \frac{\lambda^v e^{-\lambda}}{v!} = e^{-(\theta+\lambda)} \sum_{v=0}^u \frac{\theta^{u-v}}{(u-v)!} \frac{\lambda^v}{v!}, \quad u = 0, 1, 2, \dots$$

Lo anterior puede simplificarse multiplicando y dividiendo cada término entre $u!$ y usando el teorema binomial:

$$f_U(u) = \frac{e^{-(\theta+\lambda)}}{u!} \sum_{v=0}^u \binom{u}{v} \lambda^u \theta^{u-v} = \frac{e^{-(\theta+\lambda)}}{u!} (\theta + \lambda)^u, \quad u = 0, 1, 2, \dots$$

Luego $U \sim \mathcal{P}(\theta + \lambda)$.

Teorema 4.3.1 Si $X \sim \mathcal{P}(\theta)$ y $Y \sim \mathcal{P}(\lambda)$ son variables aleatorias independientes, entonces $X + Y \sim \mathcal{P}(\theta + \lambda)$.

4.3.2. Caso continuo

Si (X, Y) es un vector aleatorio continuo con función de densidad conjunta $f_{X,Y}(x, y)$, entonces la función de densidad conjunta de (U, V) puede expresarse en términos de $f_{X,Y}(x, y)$, en forma análoga a 2.1.1. Sea $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$ y $\mathcal{B} = \{(u, v) : u = g_1(x, y), v = g_2(x, y) \text{ para todo } (x, y) \in \mathcal{A}\}$. La función de densidad conjunta $f_{U,V}(u, v)$ será positiva sobre el conjunto \mathcal{B} . Si se asume que $u = g_1(x, y)$ y $v = g_2(x, y)$ definen transformaciones uno a uno de \mathcal{A} hacia \mathcal{B} entonces dichas transformaciones serán sobreyectivas según la definición de \mathcal{B} . Entonces para todo $(u, v) \in \mathcal{B}$ existe solo un $(x, y) \in \mathcal{A}$ tal que $(u, v) = (g_1(x, y), g_2(x, y))$. Para cada transformación uno a uno y sobreyectiva, se pueden resolver las ecuaciones $u = g_1(x, y)$ y $v = g_2(x, y)$ para x y y en términos de u y v . Denotemos estas transformaciones inversas por $x = h_1(u, v)$ y $y = h_2(u, v)$.

El rol que tuvo la derivada en el caso univariado ahora lo asume una cantidad llamada el *Jacobiano de la transformación*. Esta función de (u, v) , denotada por J , es el *determinante de la matriz* de derivadas parciales. Se define por:

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial y}{\partial u} \frac{\partial x}{\partial v}$$

donde:

$$\frac{\partial x}{\partial u} = \frac{\partial h_1(u, v)}{\partial u}, \quad \frac{\partial x}{\partial v} = \frac{\partial h_1(u, v)}{\partial v}, \quad \frac{\partial y}{\partial u} = \frac{\partial h_2(u, v)}{\partial u} \quad \text{y} \quad \frac{\partial y}{\partial v} = \frac{\partial h_2(u, v)}{\partial v}$$

Se asume que J es diferente de cero sobre \mathcal{B} . Entonces la función de densidad conjunta de (U, V) está dada por:

$$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v)) |J| \quad (4.3.2)$$

donde $|J|$ es el valor absoluto de J .

Ejemplo 4.3.2 Sea $X \sim \mathcal{BE}(\alpha, \beta)$ y $Y \sim \mathcal{BE}(\alpha + \beta, \gamma)$ variables aleatorias independientes. La función de densidad conjunta de (X, Y) es:

$$f_{X,Y}(x, y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha + \beta)\Gamma(\gamma)} y^{\alpha+\beta-1} (1-y)^{\gamma-1}$$

Considere las transformaciones $U = XY$ y $V = X$. Estas transformaciones van de $\mathcal{A} = \{(x, y) : 0 < x < 1, 0 < y < 1\}$ al conjunto $\mathcal{B} = \{(u, v) : 0 < u < v < 1\}$ y son uno a uno. Para todo $(u, v) \in \mathcal{B}$, las ecuaciones $u = xy$ y $v = x$ puede resolverse unicamente para $x = h_1(u, v) = v$ y $y = h_2(u, v) = u/v$.

$$J = \begin{vmatrix} 0 & 1 \\ \frac{1}{v} & -\frac{u}{v^2} \end{vmatrix} = -\frac{1}{v}$$

Luego por 4.3.2 la función de densidad conjunta es:

$$f_{U,V}(u, v) = \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} v^{\alpha-1} (1-v)^{\beta-1} \left(\frac{u}{v}\right)^{\alpha+\beta-1} \left(1 - \frac{u}{v}\right)^{\gamma-1} \frac{1}{v}$$

para $0 < u < v < 1$. La distribución marginal de $V = X$ es $\mathcal{BE}(\alpha, \beta)$. La distribución de U también lo es:

$$\begin{aligned} f_U(u) &= \int_u^1 f_{U,V}(u, v) dv \\ &= \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1} \int_u^1 \left(\frac{u}{v} - u\right)^{\beta-1} \left(1 - \frac{u}{v}\right)^{\gamma-1} \left(\frac{u}{v^2}\right) dv \end{aligned}$$

Sea $w = (u/v - u)/(1 - u)$ tal que $dw = -u/[v^2(1 - u)]dv$

$$\begin{aligned} f_U(u) &= \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1} (1 - u)^{\beta+\gamma-1} \int_0^1 w^{\beta-1} (1 - w)^{\gamma-1} dw \\ &= \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1} (1 - u)^{\beta+\gamma-1} \frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta + \gamma)} \\ &= \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta + \gamma)} u^{\alpha-1} (1 - u)^{\beta+\gamma-1}, \quad 0 < u < 1 \end{aligned}$$

Luego $U \sim \mathcal{BE}(\alpha, \beta + \gamma)$.

Ejemplo 4.3.3 Sean X y Y variables aleatorias independientes con distribución Z . Hallar $f(u, v)$ si $U = X + Y$, $V = X - Y$ y probar que son variables aleatorias independientes.

Teorema 4.3.2 Sean X y Y variables aleatorias independientes. Sea $g(x)$ una función que depende sólo de x y $h(y)$ una función que sólo depende de y . Entonces las variables aleatorias $U = g(X)$ y $V = h(Y)$ son independientes.

Prueba: Asumiendo que U y V son variables aleatorias continuas. Para todo $u \in \Re$ y $v \in \Re$ se define:

$$\mathcal{A}_u = \{x : g(x) \leq u\} \text{ y } \mathcal{B}_v = \{y : h(y) \leq v\}$$

La función de distribución acumulada conjunta de (U, V) es:

$$\begin{aligned} F_{U,V}(u, v) &= \Pr(U \leq u, V \leq v) \\ &= \Pr(X \in \mathcal{A}_u, Y \in \mathcal{B}_v) \\ &= \Pr(X \in \mathcal{A}_u) \Pr(Y \in \mathcal{B}_v) \end{aligned}$$

La función de densidad conjunta de (U, V) es:

$$f_{U,V}(u, v) = \frac{\partial^2}{\partial u \partial v} F_{U,V}(u, v) = \left(\frac{d}{du} \Pr(X \in \mathcal{A}_u) \right) \left(\frac{d}{dv} \Pr(Y \in \mathcal{B}_v) \right)$$

Luego U y V son independientes.

En muchas situaciones las transformaciones de interés no son uno a uno. Sea $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$. Suponga que A_0, A_1, \dots, A_k forma una

partición de \mathcal{A} . El conjunto A_0 , que podría ser vacío, satisface $\Pr((X, Y) \in A_0) = 0$. Las transformaciones $U = g_1(X, Y)$ y $V = g_2(X, Y)$ son uno a uno, desde A_i hacia \mathcal{B} para cada $i = 1, 2, \dots, k$. Entonces para cada i se pueden hallar las funciones inversas desde \mathcal{B} hacia A_i . Si se denotan las i -ésimas inversas por $x = h_{1i}(u, v)$ e $y = h_{2i}(u, v)$, éstas dan para $(u, v) \in \mathcal{B}$ un único $(x, y) \in A_i$ tal que $(u, v) = (g_1(x, y), g_2(x, y))$. Sea J_i el Jacobiano calculado a partir de las i -ésimas inversas. Entonces se tiene la siguiente representación de la función de densidad conjunta $f_{U,V}(u, v)$:

$$f_{U,V}(u, v) = \sum_{i=1}^k f_{X,Y}(h_{1i}(u, v), h_{2i}(u, v)) |J_i| \quad (4.3.3)$$

Ejemplo 4.3.4 Sean X y Y variables aleatorias independientes con distribución Z . Considere las transformaciones $U = X/Y$ y $V = |Y|$. Sean:

$$A_1 = \{(x, y) : y > 0\}, A_2 = \{(x, y) : y < 0\} \text{ y } A_0 = \{(x, y) : y = 0\}$$

A_0, A_1 y A_2 forman una partición de $\mathcal{A} = \mathfrak{R}^2$ y $\Pr((X, Y) \in A_0) = \Pr(Y = 0) = 0$. Para cada A_1 o A_2 , si $(x, y) \in A_i$, $v = |y| > 0$ y $u = x/y \in \mathfrak{R}$. Luego, $\mathcal{B} = \{(u, v) : v > 0\}$ es la imagen de A_1 y A_2 bajo la transformación. Además las transformaciones inversas desde \mathcal{B} hacia A_1 y desde \mathcal{B} hacia A_2 son: $x = h_{11}(u, v) = uv$, $y = h_{21}(u, v) = v$, y $x = h_{12}(u, v) = -uv$, $y = h_{22}(u, v) = -v$. Luego:

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2}$$

Usando 4.3.3

$$\begin{aligned} f_{U,V}(u, v) &= \frac{1}{2\pi} e^{-(uv)^2/2} e^{-v^2/2} |v| + \frac{1}{2\pi} e^{-(-uv)^2/2} e^{-(-v)^2/2} |v| \\ &= \frac{v}{\pi} e^{-(u^2+1)v^2/2} \end{aligned}$$

para $-\infty < u < \infty$ y $0 < v < \infty$. La distribución marginal de U es:

$$\begin{aligned}
 f_U(u) &= \int_0^\infty \frac{v}{\pi} e^{-(u^2+1)v^2/2} dv \\
 &= \frac{1}{2\pi} \int_0^\infty e^{-(u^2+1)z/2} dz \\
 &= \frac{1}{2\pi} \frac{2}{(u^2+1)} \\
 &= \frac{1}{\pi(u^2+1)}
 \end{aligned}$$

para $-\infty < u < \infty$, entonces $U \sim \mathcal{C}(0, 1)$.

4.4. Modelos jerárquicos y distribuciones mixtas

Ejemplo 4.4.1 Un insecto pone un número grande de huevos, cada uno con probabilidad de supervivencia p . En promedio, ¿cuántos huevos sobrevivirán? Sean X = Número de huevos sobrevivientes, y Y = Número de huevos puestos. Luego se tiene el siguiente modelo jerárquico:

$$X|Y \sim \mathcal{BI}(Y, p), \quad Y \sim \mathcal{P}(\lambda)$$

La variable de interés, X = Número de huevos sobrevivientes, tiene la siguiente distribución:

$$\begin{aligned}
 \Pr(X = x) &= \sum_{y=0}^{\infty} \Pr(X = x, Y = y) \\
 &= \sum_{y=0}^{\infty} \Pr(X = x|Y = y) \Pr(Y = y) \\
 &= \sum_{y=x}^{\infty} \left[\binom{y}{x} p^x (1-p)^{y-x} \right] \left[\frac{e^{-\lambda} \lambda^y}{y!} \right]
 \end{aligned}$$

ya que la probabilidad condicional es cero si $y < x$, entonces:

$$\begin{aligned}
 \Pr(X = x) &= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{((1-p)\lambda)^{y-x}}{(y-x)!} \\
 &= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{t=0}^{\infty} \frac{((1-p)\lambda)^t}{t!} \\
 &= \frac{(\lambda p)^x e^{-\lambda}}{x!} e^{(1-p)\lambda} \\
 &= \frac{(\lambda p)^x e^{-\lambda p}}{x!}
 \end{aligned}$$

es decir que $X \sim \mathcal{P}(\lambda p)$. Luego, todo proceso de inferencia marginal sobre X se hace a través de la distribución de Poisson, en la que Y no es parte del proceso. La respuesta a la pregunta inicial es $E[X] = \lambda p$, es decir, en promedio, sobrevivirán λp huevos.

Teorema 4.4.1 Si X y Y son variables aleatorias, entonces:

$$E[X] = E[E[X|Y]] \quad (4.4.1)$$

siempre que los esperados existan.

Prueba: Sea $f(x, y)$ la función de densidad de X y Y . Por definición se tiene:

$$\begin{aligned}
 E[X] &= \int_y \int_x x f(x, y) dx dy \\
 &= \int_y \left[\int_x x f(x|y) dx \right] f_Y(y) dy \\
 &= \int_y E[X|Y] f_Y(y) dy \\
 &= E_Y[E_X[X|Y]]
 \end{aligned}$$

Volviendo al ejemplo anterior:

$$E[X] = E[E[X|Y]] = E[Yp] = pE[Y] = p\lambda$$

Definición 4.4.1 Una variable aleatoria X se dice que tiene distribución mixta si la distribución de X depende de una cantidad que también tiene distribución.

En el ejemplo 4.4.1 la distribución $\mathcal{P}(\lambda p)$ es una distribución mixta ya que es el resultado de combinar una $\mathcal{BI}(Y, p)$ con $Y \sim \mathcal{P}(\lambda)$. En general, se puede afirmar que los modelos jerárquicos dan lugar a distribuciones mixtas.

Ejemplo 4.4.2 Considere una generalización del ejemplo 4.4.1:

$$X|Y \sim \mathcal{BI}(Y, p), \quad Y|\Lambda \sim \mathcal{P}(\Lambda) \quad \Lambda \sim \mathcal{E}(\beta)$$

Entonces:

$$E[X] = E[E[X|Y]] = E[pY] = pE[E[Y|\Lambda]] = pE[\Lambda] = p\beta$$

Notar que este modelo de tres niveles puede ser expresado como uno de dos jerarquías combinando los últimos dos estados. Si $Y|\Lambda \sim \mathcal{P}(\Lambda)$ y $\Lambda \sim \mathcal{E}(\beta)$, luego:

$$\begin{aligned} \Pr(Y = y) &= \int_0^\infty f(y, \lambda) d\lambda \\ &= \int_0^\infty f(y|\lambda) f(\lambda) d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda} \lambda^y}{y!} \frac{1}{\beta} e^{-\lambda/\beta} d\lambda \\ &= \frac{1}{\beta y!} \int_0^\infty \lambda^y e^{-\lambda(1+\beta^{-1})} d\lambda \\ &= \frac{1}{\beta y!} \Gamma(y+1) \left(\frac{1}{1+\beta^{-1}} \right)^{y+1} \\ &= \frac{1}{1+\beta} \left(\frac{\beta}{1+\beta} \right)^y \end{aligned}$$

La expresión para la función de probabilidad de Y es de la familia 3.1.5, es decir, binomial negativa. Luego, la jerarquía de tres estados de este ejemplo es equivalente a la jerarquía de dos estados:

$$X|Y \sim \mathcal{BI}(Y, p) \quad Y \sim \mathcal{BN}(p = \frac{1}{1+\beta}, r = 1)$$

Una mixtura útil es la Poisson-gamma, que es una generalización de una parte del modelo anterior. Si se toma la jerarquía:

$$Y|\Lambda \sim \mathcal{P}(\Lambda), \quad \Lambda \sim \mathcal{G}(\alpha, \beta)$$

entonces $Y \sim \mathcal{BN}$.

Los modelos jerárquicos pueden hacer más sencillos los cálculos. Por ejemplo, una distribución que aparece con frecuencia en estadística es la *distribución chi-cuadrado no central* con p grados de libertad y parámetro de no centralidad λ cuya función de densidad es:

$$f(X|\lambda, p) = \sum_{k=0}^{\infty} \frac{x^{p/2+k-1} e^{-x/2}}{\Gamma(p/2 + k) 2^{p/2+k}} \frac{\lambda^k e^{-\lambda}}{k!} \quad (4.4.2)$$

Calcular el $E[X]$ puede ser tarea complicada. Sin embargo, si se examina detenidamente la función de densidad puede observarse que se trata de una distribución mixta, obtenida de una densidad chi-cuadrado central y una distribución de Poisson. Es decir:

$$X|K \sim \chi_{p+2K}^2, \quad K \sim \mathcal{P}(\lambda)$$

entonces la distribución marginal de X está dada por 4.4.2. Luego:

$$E[X] = E[E[X|K]] = E[p + 2K] = p + 2\lambda$$

un cálculo bastante directo. La variancia puede ser calculada de la misma forma.

Ejemplo 4.4.3 Una generalización para ensayos de Bernoulli considera que la probabilidad de éxito no sea constante de ensayo a ensayo, manteniendo los ensayos independientes. Un modelo estándar para esta situación es:

$$X_i|p_i \sim \mathcal{B}(p_i), \quad i = 1, 2, \dots, n \quad p_i \sim \mathcal{BE}(\alpha, \beta)$$

Este modelo puede ser apropiado, por ejemplo, si se mide el éxito de una droga en n pacientes y debido a que cada paciente es diferente no es posible asumir que la probabilidad de éxito sea constante. Una variable aleatoria de interés es $Y = \sum_{i=1}^n X_i$, el número de éxitos, cuya media es:

$$E[Y] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n E[E[X_i|p_i]] = \sum_{i=1}^n E[p_i] = \sum_{i=1}^n \frac{\alpha}{\alpha + \beta} = \frac{n\alpha}{\alpha + \beta}$$

Teorema 4.4.2 Sean X y Y dos variables aleatorias, entonces:

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y]) \quad (4.4.3)$$

Prueba: Por definición:

$$\begin{aligned}
 \text{Var}[X] &= E[(X - E[X])^2] \\
 &= E[(X - E[X|Y] + E[X|Y] - E[X])^2] \\
 &= E[(X - E[X|Y])^2] + E[(E[X|Y] - E[X])^2] \\
 &\quad + 2E[(X - E[X|Y])(E[X|Y] - E[X])] \\
 &= E[(X - E[X|Y])^2] + E[(E[X|Y] - E[X])^2] \\
 &= E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])
 \end{aligned}$$

Ejemplo 4.4.4 Para calcular $\text{Var}(Y)$ del ejemplo 4.4.3:

$$\text{Var}(Y) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$$

ya que las X_i son independientes. Usando 4.4.3:

$$\text{Var}(X_i) = \text{Var}(E[X_i|p_i]) + E[\text{Var}(X_i|p_i)]$$

como $E(X_i|p_i) = p_i$ y $p_i \sim \mathcal{BE}(\alpha, \beta)$:

$$\text{Var}(E[X_i|p_i]) = \text{Var}(p_i) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

También, $X_i|p_i \sim \mathcal{B}(p_i)$, $\text{Var}(X_i|p_i) = p_i(1 - p_i)$:

$$\begin{aligned}
 E[\text{Var}(X_i|p_i)] &= E[p_i(1 - p_i)] \\
 &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 p_i(1 - p_i)p_i^{\alpha-1}(1 - p_i)^{\beta-1} dp_i \\
 &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + 1)\Gamma(\beta + 1)}{\Gamma(\alpha + \beta + 2)} \\
 &= \frac{\alpha\beta}{(\alpha + \beta)(\alpha + \beta + 1)}
 \end{aligned}$$

Como $\text{Var}(X_i)$ no depende de i , entonces:

$$\text{Var}(Y) = \frac{n\alpha\beta}{(\alpha + \beta)^2}$$

4.5. Covarianza y correlación

Definición 4.5.1 La *covarianza* de X y Y es el número definido por:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Definición 4.5.2 La *correlación* de X y Y es el número definido por:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

El valor ρ_{XY} también es llamado *coeficiente de correlación*.

Teorema 4.5.1 Sean X y Y dos variables aleatorias, entonces:

$$\text{Cov}(X, Y) = E[XY] - \mu_X \mu_Y$$

Ejemplo 4.5.1 Sea la función de densidad conjunta:

$$f(x, y) = 1 \quad 0 < x < 1 \quad x < y < x + 1$$

La distribución marginal de X es $\mathcal{U}(0, 1)$, luego $\mu_X = 1/2$ y $\sigma_X^2 = 1/12$. La distribución marginal de Y es:

$$f_Y(y) = \begin{cases} y & 0 < y < 1 \\ 2 - y & 1 \leq y < 2 \end{cases}$$

con $\mu_Y = 1$ y $\sigma_Y^2 = 1/6$. Además:

$$E[XY] = \int_0^1 \int_x^{x+1} xy dy dx = \frac{7}{12}$$

Luego $\text{Cov}(X, Y) = \frac{7}{12} - (\frac{1}{2})(1) = \frac{1}{12}$. El coeficiente de correlación es:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{1/12}{\sqrt{1/12} \sqrt{1/6}} = \frac{1}{\sqrt{2}}$$

Teorema 4.5.2 Si X y Y son variables aleatorias independientes, entonces $\text{Cov}(X, Y) = 0$ y $\rho_{XY} = 0$.

Teorema 4.5.3 Si X y Y son variables aleatorias, a y b son constantes, entonces:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

Si X y Y son variables aleatorias independientes, entonces:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

Teorema 4.5.4 Para dos variables aleatorias X y Y :

- a. $-1 \leq \rho_{XY} \leq 1$
- b. $|\rho_{XY}| = 1$ sí y solo si existen números $a \neq 0$ y b tales que $\Pr(Y = aX + b) = 1$. Si $\rho_{XY} = 1$ entonces $a > 0$ y si $\rho_{XY} = -1$ entonces $a < 0$.

Ejemplo 4.5.2 Sean $X \sim \mathcal{U}(0, 1)$ y $Z \sim \mathcal{U}(0, 1/10)$ variables aleatorias independientes. Sea $Y = X + Z$ y considere el vector aleatorio (X, Y) . La distribución conjunta de (X, Y) puede obtenerse a partir de la función de densidad conjunta de (X, Z) usando las técnicas de la sección 4.3. La función de densidad conjunta de (X, Y) es:

$$f(x, y) = 10 \quad 0 < x < 1 \quad x < y < x + \frac{1}{10}$$

El resultado anterior también puede justificarse como sigue. Dados $X = x$, $Y = x + Z$, la distribución condicional de Z dado $X = x$ es $\mathcal{U}(0, \frac{1}{10})$ ya que X y Z son independientes. Así x sirve como parámetro de locación en la distribución condicional de Y dado $X = x$, y esta distribución marginal es $\mathcal{U}(x, x + \frac{1}{10})$. Multiplicando esta función de densidad condicional por la función de densidad marginal de X , que es $\mathcal{U}(0, 1)$, se obtiene la función de densidad conjunta anterior.

Esta representación $Y = X + Z$ hace fácil el cálculo de la covarianza y el coeficiente de correlación :

$$\mathbb{E}[X] = \frac{1}{2} \quad \mathbb{E}[Y] = \mathbb{E}[X + Z] = \mathbb{E}[X] + \mathbb{E}[Z] = \frac{1}{2} + \frac{1}{20} = \frac{11}{20}$$

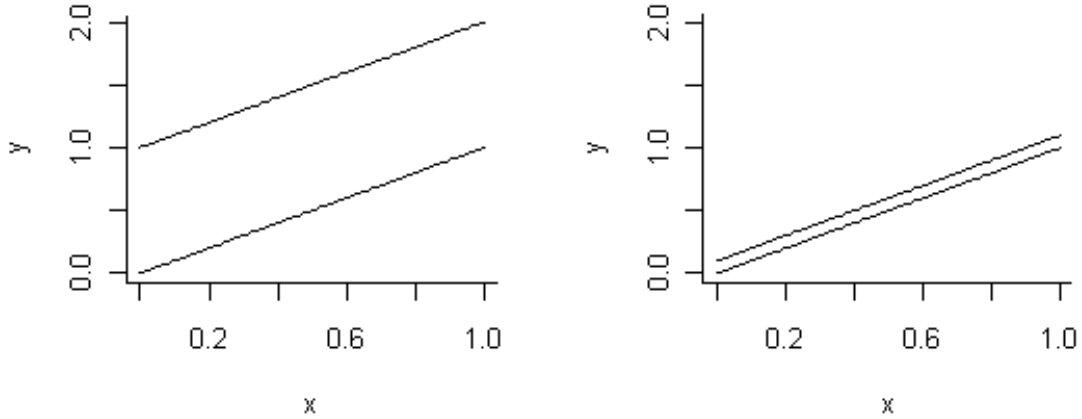
$$\begin{aligned}
\text{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\
&= E[X(X + Z)] - E[X]E[X + Z] \\
&= E[X^2] + E[XZ] - (E[X])^2 - E[X]E[Z] \\
&= \sigma_X^2 = \frac{1}{12}
\end{aligned}$$

Además $\sigma_Y^2 = \text{Var}(X + Z) = \text{Var}(X) + \text{Var}(Z) = \frac{1}{12} + \frac{1}{1200}$. Luego:

$$\rho_{XY} = \frac{\frac{1}{12}}{\sqrt{\frac{1}{12}}\sqrt{\frac{1}{12} + \frac{1}{1200}}} = \sqrt{\frac{100}{101}}$$

que es mucho mayor que $\rho_{XY} = 1/\sqrt{2}$ del ejemplo 4.5.1. ¿Por qué?

Figura 4.2: Regiones para las que $f(x, y) > 0$ de los ejemplos 4.5.1 y 4.5.2

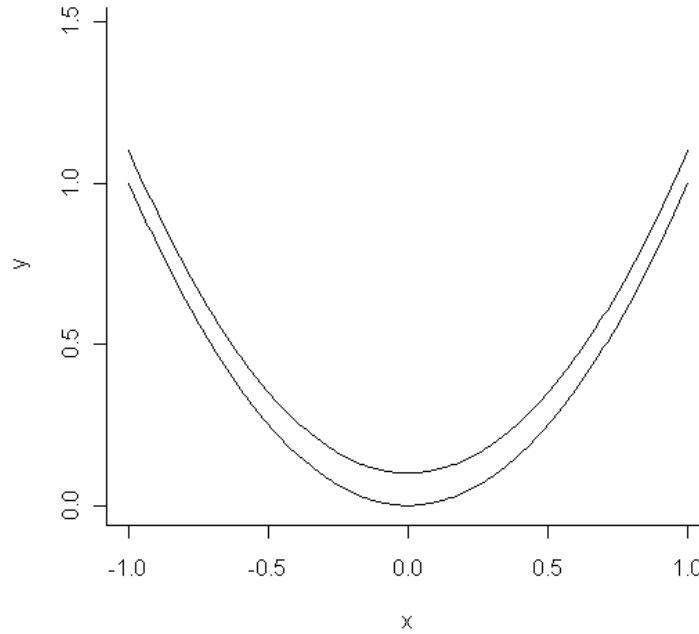


Ejemplo 4.5.3 Sean $X \sim \mathcal{U}(-1, 1)$ y $Z \sim \mathcal{U}(0, 1/10)$ variables aleatorias independientes. Sea $Y = X^2 + Z$ y considere el vector aleatorio (X, Y) . Tal como en el ejemplo 4.5.2, dado $X = x$, $Y = x^2 + Z$ y la distribución condicional de Y dado $X = x$ es $\mathcal{U}(x^2, x^2 + \frac{1}{10})$. La función de densidad conjunta de X y Y , el producto de esta condicional y la función de densidad marginal de X , es:

$$f(x, y) = 5 \quad -1 < x < 1 \quad x^2 < y < x^2 + \frac{1}{10}$$

El conjunto sobre el que $f(x, y) > 0$ se muestra en la figura 4.5.3. Existe una fuerte relación entre X y Y , tal como lo indica la distribución condicional de Y dado $X = x$, pero la relación no es lineal.

Figura 4.3: Región donde $f(x, y) > 0$ del ejemplo 4.5.3



Como $X \sim \mathcal{U}(-1, 1)$ entonces $E[X] = E[X^3] = 0$ y como X y Z son independientes, $E[XZ] = E[X]E[Z]$. Luego:

$$\begin{aligned} \text{Cov}(X, Y) &= E[X(X^2 + Z)] - E[X]E[X^2 + Z] \\ &= E[X^3] + E[XZ] - E[X]E[X^2 + Z] \\ &= 0 \end{aligned}$$

y $\rho_{X,Y} = 0$.

Definición 4.5.3 Sean $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, $0 < \sigma_X$, $0 < \sigma_Y$, y $-1 < \rho < 1$ números reales. La *función de densidad normal bivariada con medias μ_X , μ_Y varianzas σ_X^2 , σ_Y^2 y coeficiente de correlación ρ* esta dada por:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \left(\frac{x-\mu_X}{\sigma_X} \right) \left(\frac{y-\mu_Y}{\sigma_Y} \right) + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\}$$

para $-\infty < x < \infty, -\infty < y < \infty$. Algunas de las propiedades de la función de densidad conjunta anterior son:

- a. La distribución marginal de X es $\mathcal{N}(\mu_X, \sigma_X^2)$.
- b. La distribución marginal de Y es $\mathcal{N}(\mu_Y, \sigma_Y^2)$.
- c. El coeficiente de correlación entre X y Y es $\rho_{X,Y} = \rho$.
- d. Para a y b constantes, la distribución de $aX + bY$ es:

$$\mathcal{N}(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)$$

- e. Todas las distribuciones condicionales también son normales. Por ejemplo:

$$f(Y/X = x) \sim \mathcal{N}\left(\mu_Y + \rho\left(\frac{\sigma_Y}{\sigma_X}\right)(x - \mu_X), \sigma_Y^2(1 - \rho^2)\right)$$

4.6. Distribuciones multivariadas

El vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ tiene un espacio muestral que es subconjunto de \mathbb{R}^n . Si (X_1, \dots, X_n) es un vector aleatorio discreto (el espacio muestral es numerable) entonces la *función de probabilidad conjunta* de (X_1, \dots, X_n) es la función definida por $f(\mathbf{x}) = f(x_1, \dots, x_n) = \Pr(X_1 = x_1, \dots, X_n = x_n)$ para cada $(x_1, \dots, x_n) \in \mathbb{R}^n$. Entonces para todo $A \subset \mathbb{R}^n$,

$$\Pr(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A} f(\mathbf{x}) \quad (4.6.1)$$

si (X_1, \dots, X_n) es un vector aleatorio continuo, entonces la *función de densidad conjunta* de (x_1, \dots, x_n) es la función $f(x_1, \dots, x_n)$ que satisface:

$$\Pr(\mathbf{X} \in A) = \int \cdots \int_A f(\mathbf{x}) d\mathbf{x} = \int \cdots \int_A f(x_1, \cdots, x_n) dx_1 \cdots dx_n \quad (4.6.2)$$

Sea $g(\mathbf{x}) = g(x_1, \cdots, x_n)$ una función de valor real definida sobre el espacio muestral de \mathbf{X} . Entonces $g(\mathbf{X})$ es una variable aleatoria y su *valor esperado* es:

$$E[g(\mathbf{X})] = \sum_{\mathbf{x} \in \mathbb{R}^n} g(\mathbf{x}) f(\mathbf{x}) \quad \text{y} \quad E[g(\mathbf{X})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \quad (4.6.3)$$

en el caso discreto y continuo respectivamente.

La *función de probabilidad marginal* o *función de densidad marginal* para algún subconjunto de coordenadas de (X_1, \cdots, X_n) puede calcularse sumando o integrando la función de probabilidad o función de densidad conjunta sobre las otras coordenadas. Por ejemplo, la distribución marginal de (X_1, \cdots, X_k) , las primeras k -coordenadas de (X_1, \cdots, X_n) , está dada por la función de probabilidad o función de densidad:

$$f(x_1, \cdots, x_k) = \sum_{(x_{k+1}, \cdots, x_n) \in \mathbb{R}^{n-k}} f(x_1, \cdots, x_n) \quad (4.6.4)$$

$$f(x_1, \cdots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_{k+1} \cdots dx_n \quad (4.6.5)$$

para todo $(x_1, \cdots, x_k) \in \mathbb{R}^k$. La *función de probabilidad* o *función de densidad condicional* de un subconjunto de coordenadas de (x_1, \cdots, x_n) , dados los valores de las coordenadas restantes, se obtiene dividiendo la función de probabilidad o función de densidad conjunta por la función de probabilidad o función de densidad marginal de las coordenadas restantes. Así, por ejemplo, si $f(x_1, \cdots, x_n) > 0$, la función de probabilidad o función de densidad condicional de (x_{k+1}, \cdots, x_n) dados $X_1 = x_1, \cdots, X_k = x_k$ es la función de (x_{k+1}, \cdots, x_n) definida por:

$$f(x_{k+1}, \cdots, x_n | x_1, \cdots, x_k) = \frac{f(x_1, \cdots, x_n)}{f(x_1, \cdots, x_k)} \quad (4.6.6)$$

Ejemplo 4.6.1 Sea la función de densidad conjunta:

$$f(x_1, x_2, x_3, x_4) = \begin{cases} \frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2) & 0 < x_i < 1 \\ 0 & \text{de otro modo} \end{cases}$$

del vector aleatorio (X_1, \dots, X_4) . Se puede obtener la función de densidad marginal de (X_1, X_2) integrando las variables X_3 y X_4 :

$$\begin{aligned} f(x_1, x_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, \dots, x_4) dx_3 dx_4 \\ &= \int_0^1 \int_0^1 \frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2) dx_3 dx_4 \\ &= \frac{3}{4}(x_1^2 + x_2^2) + \frac{1}{2} \end{aligned}$$

para $0 < x_1 < 1, 0 < x_2 < 1$. Cualquier probabilidad o valor esperado que incluya solo X_1 y X_2 puede calcularse usando esta función de distribución marginal. Por ejemplo:

$$\begin{aligned} E[X_1 X_2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2) dx_1 dx_2 \\ &= \int_0^1 \int_0^1 x_1 x_2 \left(\frac{3}{4}(x_1^2 + x_2^2) + \frac{1}{2} \right) dx_1 dx_2 \\ &= \frac{5}{16} \end{aligned}$$

Para todo (x_1, x_2) con $0 < x_1 < 1, 0 < x_2 < 1, f(x_1, x_2) > 0$ y la función de densidad condicional de (X_3, X_4) dados $X_1 = x_1$ y $X_2 = x_2$ puede obtenerse usando 4.6.6:

$$\begin{aligned} f(x_3, x_4 | x_1, x_2) &= \frac{f(x_1, x_2, x_3, x_4)}{f(x_1, x_2)} \\ &= \frac{\frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2)}{\frac{3}{4}(x_1^2 + x_2^2) + \frac{1}{2}} \\ &= \frac{x_1^2 + x_2^2 + x_3^2 + x_4^2}{x_1^2 + x_2^2 + \frac{2}{3}} \end{aligned}$$

Definición 4.6.1 Sean n y m enteros positivos y sean p_1, \dots, p_n números tales que $0 \leq p_i \leq 1, i = 1, \dots, n$ y $\sum_{i=1}^n p_i = 1$. Entonces el vector aleatorio (X_1, \dots, X_n) tiene *distribución multinomial con m ensayos y probabilidades de celda p_1, \dots, p_n* si su función de probabilidad conjunta es:

$$f(x_1, \dots, x_n) = \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n} = m! \prod_{i=1}^n \frac{p_i^{x_i}}{x_i!}$$

sobre el conjunto de (x_1, \dots, x_n) tal que cada x_i es un entero no negativo y $\sum_{i=1}^n x_i = m$. El factor $m!/(x_1! \cdots x_n!)$ es llamado *coeficiente multinomial*.

Ejemplo 4.6.2 Considere el experimento aleatorio que consiste en lanzar 10 veces un dado. Suponga que el dado no se encuentra balanceado, tal que la probabilidad de observar i es $\frac{i}{21}$. Sea el vector aleatorio (X_1, \dots, X_6) tal que X_i representa el número de lanzamientos en los que se observó el número i , entonces su distribución es multinomial con $m = 10$ lanzamientos, $n = 6$ posibles resultados y probabilidades de celda $p_1 = \frac{1}{21}, p_2 = \frac{2}{21}, \dots, p_6 = \frac{6}{21}$. La fórmula anterior puede usarse para calcular la probabilidad de obtener el número 6 en cuatro lanzamientos, el número 5 en tres lanzamientos, el número 4 en dos lanzamientos y el número 3 en solo un lanzamiento:

$$\begin{aligned} f(0, 0, 1, 2, 3, 4) &= \frac{10!}{0!0!1!2!3!4!} \left(\frac{1}{21}\right)^0 \left(\frac{2}{21}\right)^0 \left(\frac{3}{21}\right)^1 \left(\frac{4}{21}\right)^2 \left(\frac{5}{21}\right)^3 \left(\frac{6}{21}\right)^4 \\ &= \frac{59}{10000} \end{aligned}$$

Teorema 4.6.1 (Teorema Multinomial) Sean m y n enteros positivos. Sea \mathcal{A} el conjunto de vectores $x = (x_1, \dots, x_n)$ tal que cada x_i es un entero no negativo y $\sum_{i=1}^n x_i = m$. Entonces, para números reales p_1, p_2, \dots, p_n :

$$(p_1 + \cdots + p_n)^m = \sum_{\mathbf{x} \in \mathcal{A}} \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n}$$

Definición 4.6.2 Sean $\mathbf{X}_1, \dots, \mathbf{X}_n$ vectores aleatorios con función de probabilidad o función de densidad conjunta $f(\mathbf{x}_1, \dots, \mathbf{x}_n)$. Sea $f_{\mathbf{X}_i}(\mathbf{x}_i)$ la función de probabilidad o función de densidad marginal de \mathbf{X}_i . Entonces $\mathbf{X}_1, \dots, \mathbf{X}_n$ son *vectores aleatorios mutuamente independientes* si, para todo $(\mathbf{x}_1, \dots, \mathbf{x}_n)$:

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n) = f_{\mathbf{X}_1}(\mathbf{x}_1) \cdots f_{\mathbf{X}_n}(\mathbf{x}_n) = \prod_{i=1}^n f_{\mathbf{X}_i}(\mathbf{x}_i)$$

si todas las \mathbf{X}_i 's son unidimensionales, entonces X_1, \dots, X_n son llamadas *variables aleatorias mutuamente independientes*.

Teorema 4.6.2 Sean X_1, \dots, X_n variables aleatorias mutuamente independientes. Sean g_1, \dots, g_n funciones de valor real tales que $g_i(x_i)$ es una función solo de x_i , $i = 1, 2, \dots, n$, entonces:

$$E[g_1(X_1) \cdots g_n(X_n)] = E[g_1(X_1)] \cdots E[g_n(X_n)]$$

Teorema 4.6.3 (Generalización del teorema 4.2.2) Sean X_1, \dots, X_n variables aleatorias mutuamente independientes con funciones generatrices de momentos $M_{X_1}(t), \dots, M_{X_n}(t)$. Si $Z = X_1 + \dots + X_n$, entonces la función generatriz de momentos de Z es:

$$M_Z(t) = M_{X_1}(t) \cdots M_{X_n}(t)$$

En particular, si las variables aleatorias X_1, \dots, X_n tienen la misma distribución, con función generatriz de momentos $M_X(t)$, entonces:

$$M_Z(t) = [M_X(t)]^n$$

Ejemplo 4.6.3 Suponga X_1, \dots, X_n son variables aleatorias mutuamente independientes y la distribución de $X_i \sim \mathcal{G}(\alpha_i, \beta)$. La función generatriz de momentos de una $\mathcal{G}(\alpha, \beta)$ es $M_X(t) = (1 - \beta t)^{-\alpha}$. Si $Z = X_1 + \dots + X_n$ su función generatriz de momentos es:

$$\begin{aligned} M_Z(t) &= M_{X_1}(t) \cdots M_{X_n}(t) \\ &= (1 - \beta t)^{-\alpha_1} \cdots (1 - \beta t)^{-\alpha_n} \\ &= (1 - \beta t)^{-(\alpha_1 + \cdots + \alpha_n)} \end{aligned}$$

entonces $Z \sim \mathcal{G}(\alpha_1 + \cdots + \alpha_n, \beta)$.

Corolario 4.6.1 Sean X_1, \dots, X_n variables aleatorias mutuamente independientes con funciones generatrices de momentos $M_{X_1}(t), \dots, M_{X_n}(t)$. Si a_1, \dots, a_n y b_1, \dots, b_n son constantes, entonces la función generatriz de momentos de $Z = (a_1 X_1 + b_1) + \dots + (a_n X_n + b_n)$ es:

$$M_Z(t) = \left(e^{t(\sum b_i)} \right) M_{X_1}(a_1 t) \cdots M_{X_n}(a_n t)$$

Ejemplo 4.6.4 Sean X_1, \dots, X_n variables aleatorias mutuamente independientes con $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Si a_1, \dots, a_n y b_1, \dots, b_n son constantes, entonces:

$$Z = \sum_{i=1}^n (a_i X_i + b_i) \sim \mathcal{N} \left(\sum_{i=1}^n (a_i \mu_i + b_i), \sum_{i=1}^n a_i^2 \sigma_i^2 \right)$$

Teorema 4.6.4 (Generalización del teorema 4.3.2) Sean $\mathbf{X}_1, \dots, \mathbf{X}_n$ vectores aleatorios independientes. Sea $g_i(\mathbf{x}_i)$ una función que solo depende de \mathbf{x}_i , $i = 1, \dots, n$. Entonces las variables aleatorias $U_i = g_i(\mathbf{X}_i)$ son mutuamente independientes.

4.7. Transformaciones sobre un vector aleatorio

Sea $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ un vector aleatorio con función de densidad $f_{\mathbf{X}}(\mathbf{x})$. Sea $\mathcal{A} = \{\mathbf{x} : f_{\mathbf{X}}(\mathbf{x}) > 0\}$. Considere un nuevo vector aleatorio (U_1, \dots, U_n) definido por $U_i = g_i(\mathbf{X}_1, \dots, \mathbf{X}_n)$, $i = 1, \dots, n$. Suponga que A_0, A_1, \dots, A_k forman una partición de \mathcal{A} . El conjunto A_0 , el cual podría ser vacío, satisface $\Pr((\mathbf{X}_1, \dots, \mathbf{X}_n) \in A_0) = 0$. La transformación (U_1, \dots, U_n) es una transformación 1 - 1 desde A_i hacia \mathcal{B} . Entonces para cada i , se puede obtener las funciones inversas desde \mathcal{B} hacia A_i .

Denote la i -ésima inversa por $x_1 = h_{1i}(u_1, \dots, u_n)$, $x_2 = h_{2i}(u_1, \dots, u_n), \dots, x_n = h_{ni}(u_1, \dots, u_n)$. Estas inversas dan un único $(x_1, \dots, x_n) \in A_i$ tal que $(u_1, \dots, u_n) = (g_1(x_1, \dots, x_n), \dots, g_n(x_1, \dots, x_n))$. Sea J_i el jacobiano calculado desde la i -ésima inversa, es decir:

$$J_i = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \dots & \frac{\partial x_1}{\partial u_n} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \dots & \frac{\partial x_2}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \dots & \frac{\partial x_n}{\partial u_n} \end{vmatrix} = \begin{vmatrix} \frac{\partial h_{1i}(\mathbf{u})}{\partial u_1} & \frac{\partial h_{1i}(\mathbf{u})}{\partial u_2} & \dots & \frac{\partial h_{1i}(\mathbf{u})}{\partial u_n} \\ \frac{\partial h_{2i}(\mathbf{u})}{\partial u_1} & \frac{\partial h_{2i}(\mathbf{u})}{\partial u_2} & \dots & \frac{\partial h_{2i}(\mathbf{u})}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_{ni}(\mathbf{u})}{\partial u_1} & \frac{\partial h_{ni}(\mathbf{u})}{\partial u_2} & \dots & \frac{\partial h_{ni}(\mathbf{u})}{\partial u_n} \end{vmatrix}$$

es el determinante de la matriz $n \times n$. Luego la función de densidad conjunta, $f_{\mathbf{U}}(u_1, \dots, u_n)$ para $\mathbf{u} \in \mathcal{B}$ es:

$$f_{\mathbf{U}}(u_1, \dots, u_n) = \sum_{i=1}^k f_{\mathbf{X}}(h_{i1}(u_1, \dots, u_n), \dots, (h_{in}(u_1, \dots, u_n))) |J_i| \quad (4.7.1)$$

Ejemplo 4.7.1 Sea (X_1, X_2, X_3, X_4) con función de densidad conjunta:

$$f_{\mathbf{X}}(x_1, x_2, \dots, x_4) = 24e^{-x_1-x_2-x_3-x_4}, \quad 0 < x_1 < x_2 < x_3 < x_4 < \infty$$

Considere las transformaciones:

$$U_1 = X_1, U_2 = X_2 - X_1, U_3 = X_3 - X_2, U_4 = X_4 - X_3$$

que van del conjunto \mathcal{A} sobre el conjunto $\mathcal{B} = \{u : 0 < u_i < \infty\}$ para $i = 1, 2, 3, 4$. Además son transformaciones 1 - 1, luego $k = 1$ y la inversa es:

$$X_1 = U_1, X_2 = U_1 + U_2, X_3 = U_1 + U_2 + U_3, X_4 = U_1 + U_2 + U_3 + U_4$$

El Jacobiano de la inversa es:

$$J = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{vmatrix} = 1$$

Luego:

$$\begin{aligned} f_{\mathbf{U}}(u_1, \dots, u_4) &= 24e^{-u_1-(u_1+u_2)-(u_1+u_2+u_3)-(u_1+u_2+u_3+u_4)} \\ &= 24e^{-4u_1-3u_2-2u_3-u_4} \quad \text{sobre } \mathcal{B} \end{aligned}$$

Además $f_U(u_i) = (5-i)e^{-(5-i)u_i}$, $u_i > 0$, es decir $U_i \sim \mathcal{E}(1/(5-i))$. Del teorema 4.3.2 se puede observar que U_1, U_2, \dots, U_4 son variables aleatorias mutuamente independientes.

4.8. Desigualdades

4.8.1. Desigualdades numéricas

Lema 4.8.1 Sean a, b números positivos y p, q números positivos mayores que 1 que satisfacen:

$$\frac{1}{p} + \frac{1}{q} = 1 \tag{4.8.1}$$

entonces:

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab \tag{4.8.2}$$

Teorema 4.8.1 (Desigualdad de Holder) Sean X y Y son variables aleatorias que satisfacen 4.8.1, entonces:

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{1/p} (\mathbb{E}[|Y|^q])^{1/q} \quad (4.8.3)$$

Teorema 4.8.2 (Desigualdad de Cauchy-Schwarz) Si X y Y son variables aleatorias, entonces:

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^2])^{1/2} (\mathbb{E}[|Y|^2])^{1/2} \quad (4.8.4)$$

Capítulo 5

Propiedades en una muestra aleatoria

5.1. Conceptos básicos sobre muestras aleatorias

Definición 5.1.1 X_1, \dots, X_n son llamadas una *muestra aleatoria de tamaño n de una población $f(x)$* si son variables aleatorias mutuamente independientes y la función de probabilidad o densidad marginal de cada X_i es $f(x)$. Alternativamente, X_1, \dots, X_n son llamadas variables aleatorias independientes e idénticamente distribuidas con función de probabilidad o densidad $f(x)$.

Si la función de probabilidad o densidad es miembro de una familia paramétrica $f(x|\theta)$, entonces la función de probabilidad o densidad conjunta es:

$$f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) \quad (5.1.1)$$

Ejemplo 5.1.1 Sea X_1, \dots, X_n una muestra aleatoria de una población $\mathcal{E}(\beta)$ que corresponden al tiempo de funcionamiento (en años) de n circuitos idénticos sometidos a prueba. La función de densidad conjunta de la muestra es:

$$f(x_1, \dots, x_n | \beta) = \prod_{i=1}^n f(x_i | \beta) = \prod_{i=1}^n \frac{1}{\beta} e^{-x_i/\beta} = \frac{1}{\beta^n} e^{-(x_1 + \dots + x_n)/\beta}$$

La probabilidad que todos los circuitos funcionen al menos dos años es:

$$\begin{aligned} \Pr(X_1 > 2, \dots, X_n > 2) &= \int_2^\infty \dots \int_2^\infty \frac{1}{\beta^n} e^{-(x_1 + \dots + x_n)/\beta} dx_1 \dots dx_n \\ &= e^{-2/\beta} \int_2^\infty \dots \int_2^\infty \frac{1}{\beta^{n-1}} e^{-(x_2 + \dots + x_n)/\beta} dx_2 \dots dx_n \\ &= (e^{-2/\beta})^n = e^{-2n/\beta} \end{aligned}$$

Usando independencia:

$$\begin{aligned} \Pr(X_1 > 2, \dots, X_n > 2) &= \Pr(X_1 > 2) \dots \Pr(X_n > 2) \\ &= (e^{-2/\beta})^n = e^{-2n/\beta} \end{aligned}$$

Ejemplo 5.1.2 Suponga $\{1, \dots, 1000\}$ es una población finita, es decir $N = 1000$. Si se elige una muestra sin reemplazo de tamaño $n = 10$, ¿cuál es la probabilidad que los 10 valores sean mayores a 200? Si se asume que X_1, \dots, X_{10} son independientes:

$$\begin{aligned} \Pr(X_1 > 200, \dots, X_{10} > 200) &= \Pr(X_1 > 200) \dots \Pr(X_{10} > 200) \\ &= \left(\frac{800}{1000}\right)^{10} = 0,107374 \end{aligned}$$

Sea Y = número de valores en la muestra que son mayores a 200, entonces $Y \sim \mathcal{H}(N = 1000, M = 800, k = 10)$. Luego:

$$\begin{aligned} \Pr(X_1 > 200, \dots, X_{10} > 200) &= \Pr(Y = 10) \\ &= \frac{\binom{800}{10} \binom{200}{0}}{\binom{1000}{10}} = 0,106164 \end{aligned}$$

5.2. Sumas de variables aleatorias a partir de una muestra aleatoria

Definición 5.2.1 Sea X_1, \dots, X_n una muestra aleatoria de tamaño n de una población y sea $T(X_1, \dots, X_n)$ una función cuyo dominio incluye el espacio

muestral de (X_1, \dots, X_n) , entonces la variable aleatoria $Y = T(X_1, \dots, X_n)$ es llamada una *estadística* cuya distribución es llamada la *distribución de muestreo de Y*.

Definición 5.2.2 La *media muestral* es el promedio aritmético de los valores en la muestra aleatoria. Usualmente se denota por:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

Definición 5.2.3 La *varianza muestral* es la estadística definida por:

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

La *desviación estándar muestral* es la estadística definida por $S = \sqrt{S^2}$.

Teorema 5.2.1 Sean x_1, \dots, x_n números cualesquiera y $\bar{x} = (x_1 + \dots + x_n)/n$, entonces:

- a. $\min \sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2$
- b. $(n-1)s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$

Lema 5.2.1 Sea X_1, \dots, X_n una muestra aleatoria de una población y sea $g(x)$ una función tal que $E[g(X_1)]$ y $\text{Var}(g(X_1))$ existen, entonces:

$$E\left[\sum_{i=1}^n g(X_i)\right] = nE[g(X_1)] \quad \text{y} \quad \text{Var}\left(\sum_{i=1}^n g(X_i)\right) = n\text{Var}(g(X_1)) \quad (5.2.1)$$

Teorema 5.2.2 Sea X_1, \dots, X_n una muestra aleatoria de una población con media μ y varianza $\sigma^2 < \infty$, entonces:

- a. $E[\bar{X}] = \mu$
- b. $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$
- c. $E[S^2] = \sigma^2$

Teorema 5.2.3 Sea X_1, \dots, X_n una muestra aleatoria de una población con función generatriz de momentos $M_X(t)$, entonces la función generatriz de momentos de la media muestral es:

$$M_{\bar{X}}(t) = [M_X(t/n)]^n$$

Ejemplo 5.2.1 Sea X_1, \dots, X_n una muestra aleatoria de una población $\mathcal{N}(\mu, \sigma^2)$. La función generatriz de momentos de la media muestral es:

$$\begin{aligned} M_{\bar{X}}(t) &= \left[\exp \left\{ \mu \frac{t}{n} + \frac{\sigma^2 (t/n)^2}{2} \right\} \right]^n \\ &= \exp \left\{ \mu t + \frac{(\sigma^2/n)t^2}{2} \right\} \end{aligned}$$

entonces, $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.

Ejemplo 5.2.2 Sea X_1, \dots, X_n una muestra aleatoria de una población $\mathcal{G}(\alpha, \beta)$. La función generatriz de momentos de la media muestral es:

$$\begin{aligned} M_{\bar{X}}(t) &= \left[\left(\frac{1}{1 - \beta(t/n)} \right)^\alpha \right]^n \\ &= \left(\frac{1}{1 - (\beta/n)t} \right)^{n\alpha} \end{aligned}$$

entonces, $\bar{X} \sim \mathcal{G}(n\alpha, \beta/n)$.

Si el teorema 5.2.3 no es aplicable puede usarse el método de la transformación para hallar la función de densidad de $Y = (X_1 + \dots + X_n)$ y luego la de \bar{X} . En estos casos la siguiente fórmula es útil.

Teorema 5.2.4 Si X y Y son variables aleatorias continuas independientes con función de densidad $f_X(x)$ y $f_Y(y)$, entonces la función de densidad de $Z = X + Y$ es:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z - w) dw \quad (5.2.2)$$

Ejemplo 5.2.3 Suponga que desea obtener la distribución de \bar{Z} , la media de Z_1, \dots, Z_n observaciones independientes e idénticamente distribuidas según

$\mathcal{C}(0, 1)$. Aplicando 5.2.2 se puede demostrar que $Z_1 + \cdots + Z_n$ tiene distribución $\mathcal{C}(0, n)$ y $\bar{Z} \sim \mathcal{C}(0, 1)$, es decir la media muestral tiene la misma distribución de las observaciones individuales.

Suponga X_1, \dots, X_n es una muestra de $(1/\sigma)f((x-\mu)/\sigma)$ un miembro de una familia de locación-escala. Entonces la distribución de \bar{X} tiene relación con la distribución de \bar{Z} , la media muestral a partir de la función de densidad estándar $f(z)$. Recordar que según el teorema 3.4.2 existen variables aleatorias Z_1, \dots, Z_n tales que $X_i = \sigma Z_i + \mu$ y la función de densidad de cada Z_i es $f(z)$. Además se tiene que Z_1, \dots, Z_n son mutuamente independientes por lo que se trata de una muestra aleatoria de $f(z)$. Las medias muestrales \bar{X} y \bar{Z} están relacionadas por:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n (\sigma Z_i + \mu) = \sigma \bar{Z} + \mu$$

Luego, aplicando nuevamente el teorema 3.4.2, se tiene que si $g(z)$ es la función de densidad de \bar{Z} , entonces $(1/\sigma)g((x-\mu)/\sigma)$ es la función de densidad de \bar{X} .

Ejemplo 5.2.4 Si Z_1, \dots, Z_n es una muestra aleatoria de una distribución de $\mathcal{C}(0, 1)$, entonces $\bar{Z} \sim \mathcal{C}(0, 1)$. Luego si $X_1, \dots, X_n \sim \mathcal{C}(\mu, \sigma)$ entonces $\bar{X} \sim \mathcal{C}(\mu, \sigma)$. Es importante notar que la dispersión de \bar{X} medida por σ , es la misma sin tomar en cuenta el tamaño de muestra n , lo cual entra en aparente contradicción con el teorema 5.2.2 ya que la distribución de Cauchy no tiene varianza finita.

Cuando se muestrea a partir de una familia exponencial la distribución de muestreo de la suma es fácil de encontrar. Las estadísticas T_1, \dots, T_k en el siguiente teorema son medidas de resumen importantes, tal como se verá más adelante.

Teorema 5.2.5 Suponga X_1, \dots, X_n es una muestra aleatoria tomada de una función de probabilidad o densidad $f(x|\theta)$ donde:

$$f(x|\theta) = h(x)c(\theta) \exp \left\{ \sum_{i=1}^k w_i(\theta) T_i(x) \right\}$$

es un miembro de una familia exponencial. Se definen las estadísticas T_1, \dots, T_k por:

$$T_i(X_1, \dots, X_n) = \sum_{j=1}^n T_i(X_j), \quad i = 1, \dots, k.$$

Suponga que:

$$\{(w_1(\theta), \dots, w_k(\theta) : \theta \in \Theta\}$$

y

$$\{(T_1(x_1, \dots, x_n), \dots, T_k(x_1, \dots, x_n)) : x_j \in \mathcal{X}\}$$

donde $n \geq k$, Θ es el espacio paramétrico y \mathcal{X} es el espacio muestral de X_j . Luego, la distribución de (T_1, \dots, T_k) es miembro de una familia exponencial de la forma:

$$f_T(t_1, \dots, t_k | \theta) = H(t_1, \dots, t_k) [c(\theta)]^n \exp \left\{ \sum_{i=1}^k w_i(\theta) t_i \right\}$$

Ejemplo 5.2.5 Suponga que X_1, \dots, X_n es una muestra aleatoria de una distribución $\mathcal{B}(p)$ que pertenece a una familia exponencial con $k = 1$, $c(p) = (1 - p)$, $w_1(p) = \log(p/(1 - p))$ y $t_1(x) = x$. Luego $T_1 = X_1 + \dots + X_n \sim \mathcal{BI}(n, p)$ que pertenece a una familia exponencial con $w_1(p)$ y $c(p) = (1 - p)$.

5.3. Muestreo desde la distribución Normal

5.3.1. Propiedades de la media y variancia muestral

Teorema 5.3.1 Sea X_1, \dots, X_n una muestra aleatoria de la distribución $\mathcal{N}(\mu, \sigma^2)$ y sean $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ y $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Entonces:

- a. \bar{X} y S^2 son variables aleatorias independientes.
- b. $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.
- c. $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$.

Lema 5.3.1 Sea χ_p^2 una variable aleatoria con distribución chi-cuadrado con p grados de libertad.

- a. Si $Z \sim \mathcal{N}(0, 1)$ entonces $Z^2 \sim \chi_1^2$.

- b. Si X_1, \dots, X_n son independientes y $X_i \sim \chi_{p_i}^2$ entonces $X_1 + \dots + X_n \sim \chi_{p_1 + \dots + p_n}^2$.

Lema 5.3.2 Sea $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, \dots, n$ independientes. Para constantes a_{ij} y b_{rj} ($i = 1, \dots, k; r = 1, \dots, m$) donde $k + m \leq n$, se define:

$$U_i = \sum_{j=1}^n a_{ij} X_j, \quad i = 1, \dots, k$$

$$V_r = \sum_{j=1}^n b_{rj} X_j, \quad r = 1, \dots, m$$

- a. Las variables aleatorias U_i y V_r son independientes si y solo si $\text{Cov}(U_i, V_r) = \sum_{j=1}^n a_{ij} b_{rj} \sigma_j^2 = 0$.
- b. Los vectores aleatorios (U_1, \dots, U_k) y (V_1, \dots, V_m) son independientes sí y solo si U_i es independiente de V_r para todos los pares i, r ($i = 1, \dots, k; r = 1, \dots, m$).

5.3.2. Distribuciones derivadas: t de Student y F de Snedecor

Si X_1, \dots, X_n es una muestra aleatoria de la distribución $\mathcal{N}(\mu, \sigma^2)$ se sabe que:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (5.3.1)$$

tiene distribución $\mathcal{N}(0, 1)$. Muchas veces σ no se conoce, en este caso el proceso de inferencia se realiza sobre:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \quad (5.3.2)$$

Definición 5.3.1 Sea X_1, \dots, X_n es una muestra aleatoria de una distribución $\mathcal{N}(\mu, \sigma^2)$. La cantidad $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ tiene *distribución t-student con $n - 1$ grados de libertad*. Equivalentemente, una variable aleatoria T tiene distribución t de student con p grados de libertad, y se denota por $T \sim t_p$, si tiene la siguiente función de densidad:

$$f_T(t) = \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1 + t^2/p)^{(p+1)/2}}, \quad -\infty < t < \infty \quad (5.3.3)$$

Si $p = 1$ entonces 5.3.3 se convierte en la distribución Cauchy, lo cual ocurre cuando el tamaño de muestra es 2.

La distribución t de student no tiene función generatriz de momentos ya que no todos sus momentos existen. De hecho, si existen p grados de libertad, entonces solo existen $p - 1$ momentos. Por lo tanto, t_1 no tiene media, t_2 no tiene variancia, etc. Es fácil chequear que si T_p es una variable aleatoria con distribución t_p entonces:

$$\begin{aligned} E[T_p] &= 0 & \text{si } p > 1 \\ \text{Var}(T_p) &= \frac{p}{p-2} & \text{si } p > 2 \end{aligned} \quad (5.3.4)$$

Definición 5.3.2 Sea X_1, \dots, X_n una muestra aleatoria de una población $\mathcal{N}(\mu_X, \sigma_X^2)$ y sea Y_1, \dots, Y_m una muestra aleatoria de una población independiente $\mathcal{N}(\mu_Y, \sigma_Y^2)$. La variable aleatoria $F = (S_X^2/\sigma_X^2)/(S_Y^2/\sigma_Y^2)$ tiene *distribución \mathcal{F} de Snedecor con $n - 1$ y $m - 1$ grados de libertad*. Equivalentemente, la variable aleatoria F tiene distribución \mathcal{F} con p y q grados de libertad, si su función de densidad es:

$$f_F(x) = \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})} \left(\frac{p}{q}\right)^{p/2} \frac{x^{(p/2)-1}}{[1 + (p/q)x]^{(p+q)/2}}, \quad 0 < x < \infty \quad (5.3.5)$$

Teorema 5.3.2 Usando la técnica de la transformación es posible establecer los siguientes resultados:

- a. Si $X \sim \mathcal{F}_{p,q}$ entonces $1/X \sim \mathcal{F}_{q,p}$.
- b. Si $X \sim t_q$ entonces $X^2 \sim \mathcal{F}_{1,q}$.
- c. Si $X \sim \mathcal{F}_{p,q}$ entonces $(p/q)X/(1 + (p/q)X) \sim \mathcal{BE}(p/2, q/2)$.

5.4. Estadísticas de orden

Definición 5.4.1 Las *estadísticas de orden* de una muestra X_1, \dots, X_n son los valores puestos en orden ascendente y se denotan por $X_{(1)}, \dots, X_{(n)}$.

Las estadísticas de orden son variables aleatorias que satisfacen $X_{(1)} \leq \dots \leq X_{(n)}$. En particular:

$$\begin{aligned} X_{(1)} &= \min_{1 \leq i \leq n} X_i \\ X_{(2)} &= \text{segundo valor más pequeño } X_i \\ &\vdots \\ X_{(n)} &= \max_{1 \leq i \leq n} X_i \end{aligned}$$

El *rango muestral*, $R = X_{(n)} - X_{(1)}$ es la diferencia entre la observación más grande y pequeña. La *mediana muestral*, denotada por M , es el número tal que aproximadamente la mitad de las observaciones son menores que M y la otra mitad es mayor. En términos de las estadísticas de orden, M se define por:

$$M = \begin{cases} X_{((n+1)/2)} & \text{si } n \text{ es impar} \\ (X_{(n/2)} + X_{(n/2+1)})/2 & \text{si } n \text{ es par} \end{cases} \quad (5.4.1)$$

Para todo número p entre 0 y 1, el percentil muestral $100p$ es la observación tal que aproximadamente np de las observaciones son menores que el mencionado percentil y $n(1-p)$ de las observaciones restantes son mayores. El percentil muestral 50 es la mediana muestral. Para otros valores de p , se puede definir de forma precisa los percentiles muestrales en términos de las estadísticas de orden.

Definición 5.4.2 La notación $\{b\}$, cuando aparezca como subíndice, se define como *el número b redondeado al entero más cercano*, es decir, si i es un número entero y $i - 0,5 \leq b < i + 0,5$, entonces $\{b\} = i$.

El percentil $100p$ es $X_{\{np\}}$ si $\frac{1}{2n} < p < 0,5$ y $X_{(n+1-\{n(1-p)\})}$ si $0,5 < p < 1 - \frac{1}{2n}$. Por ejemplo, si $n = 12$ y se desea hallar el percentil 65, se tiene $12(1 - 0,65) = 4,2$ y $12 + 1 - 4 = 9$. Luego el percentil 65 es $X_{(9)}$.

Teorema 5.4.1 Sea X_1, \dots, X_n una muestra aleatoria de una distribución discreta con función de probabilidad $f_X(x_i) = p_i$ donde $x_1 < x_2 < \dots$ son

los posibles valores de X en orden ascendente. Se definen:

$$\begin{aligned} P_0 &= 0 \\ P_1 &= p_1 \\ P_2 &= p_1 + p_2 \\ &\vdots \\ P_i &= p_1 + p_2 + \cdots + p_i \end{aligned}$$

Sean $X_{(1)}, \dots, X_{(n)}$ las estadísticas de orden de la muestra, entonces:

$$\Pr(X_{(j)} \leq x_i) = \sum_{k=j}^n \binom{n}{k} P_i^k (1 - P_i)^{n-k} \quad (5.4.2)$$

y

$$\Pr(X_{(j)} = x_i) = \sum_{k=j}^n \binom{n}{k} \left[P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k} \right] \quad (5.4.3)$$

Teorema 5.4.2 Sean $X_{(1)}, \dots, X_{(n)}$ las estadísticas de orden de una muestra aleatoria de una población continua con función de distribución acumulada $F_X(x)$ y función de densidad $f_X(x)$. Entonces la función de densidad de $X_{(j)}$ es:

$$f_{X_{(j)}}(u) = \frac{n!}{(j-1)!(n-j)!} f_X(u) [F_X(u)]^{j-1} [1 - F_X(u)]^{n-j} \quad (5.4.4)$$

Ejemplo 5.4.1 Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas como $\mathcal{U}(0, 1)$, tal que $f_X(x) = 1$ y $F_X(x) = x$ para $x \in (0, 1)$. Usando el teorema 5.4.2 se tiene que la función de densidad de la j -ésima estadística de orden es:

$$\begin{aligned} f_{X_{(j)}}(u) &= \frac{n!}{(j-1)!(n-j)!} u^{j-1} (1-u)^{n-j} \\ &= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} u^{j-1} (1-u)^{(n-j+1)-1} \end{aligned}$$

Es decir, $X_{(j)} \sim \mathcal{BE}(j, n-j+1)$. Luego:

$$\mathbb{E}[X_{(j)}] = \frac{j}{n+1} \quad \text{y} \quad \text{Var}(X_{(j)}) = \frac{j(n-j+1)}{(n+1)^2(n+2)}$$

Teorema 5.4.3 Sean $X_{(1)}, \dots, X_{(n)}$ las estadísticas de orden de una muestra aleatoria X_1, \dots, X_n de una población continua con función de distribución acumulada $F_X(x)$ y función de densidad $f_X(x)$. Entonces la función de densidad conjunta de $X_{(i)}$ y $X_{(j)}$, $1 \leq i < j \leq n$, es:

$$f_{X_{(i)}, X_{(j)}}(u, v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) [F_X(u)]^{i-1} [F_X(v) - F_X(u)]^{j-1-i} [1 - F_X(v)]^{n-j} \quad (5.4.5)$$

para $-\infty < u < v < \infty$.

Es posible obtener la función de densidad conjunta de tres o más estadísticas de orden usando argumentos similares pero más complicados. La función de densidad conjunta de todas las estadísticas de orden está dada por:

$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = \begin{cases} n! f_X(x_1) \cdots f_X(x_n) & -\infty < x_1 < \cdots < x_n < \infty \\ 0 & \text{de otro modo} \end{cases}$$

Ejemplo 5.4.2 Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas según $\mathcal{U}(0, a)$ y sean $X_{(1)}, \dots, X_{(n)}$ las estadísticas de orden. El rango muestral fue definido como $R = X_{(n)} - X_{(1)}$. El *rango medio* o *semirango* es una medida de localización como la mediana o media muestral, y se define por $S = (X_{(1)} + X_{(n)})/2$. Se pide hallar la función de densidad conjunta de R y S a partir de la función de densidad conjunta de $X_{(1)}$ y $X_{(n)}$. De 5.4.5 se tiene que:

$$f_{X_{(1)}, X_{(n)}}(u, v) = \frac{n(n-1)(v-u)^{n-2}}{a^n}, \quad 0 < u < v < a$$

Resolviendo para $X_{(1)}$ y $X_{(n)}$ se obtiene:

$$X_{(1)} = S - R/2$$

$$X_{(n)} = S + R/2$$

Las transformaciones anteriores son 1-1 y van desde $\mathcal{A} = \{(u, v) : 0 < u < v < a\}$ hacia $\mathcal{B} = \{(r, s) : 0 < r < a, r/2 < s < a - r/2\}$. El jacobiano es:

$$J = \begin{vmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{vmatrix} = 1$$

La función de densidad conjunta de (R, S) es:

$$f_{R,S}(r, s) = \frac{n(n-1)r^{n-2}}{a^n}, \quad 0 < r < a, \quad r/2 < s < a - r/2$$

La función de densidad marginal de R es:

$$f_R(r) = \int_{r/2}^{a-r/2} \frac{n(n-1)r^{n-2}}{a^n} ds = \frac{n(n-1)r^{n-2}(a-r)}{a^n} \quad (5.4.6)$$

Si $a = 1$, se tiene que $r \sim \mathcal{BE}(n-1, 2)$. Se puede probar que R/a tiene distribución beta, donde a es el parámetro de escala. La función de densidad marginal de S es:

$$f_S(s) = \int_0^{2s} \frac{n(n-1)r^{n-2}}{a^n} dr = \frac{n(2s)^{n-1}}{a^n}, \quad 0 < s \leq a/2$$

y

$$f_S(s) = \int_0^{2(a-s)} \frac{n(n-1)r^{n-2}}{a^n} dr = \frac{n[2(a-s)]^{n-1}}{a^n}, \quad a/2 < s \leq a$$

5.5. Conceptos de convergencia

5.5.1. Convergencia en probabilidad

Definición 5.5.1 Una secuencia de variables aleatorias X_1, X_2, \dots converge en probabilidad hacia la variable aleatoria X , si para todo $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| \geq \epsilon) = 0 \quad \text{ó} \quad \lim_{n \rightarrow \infty} \Pr(|X_n - X| < \epsilon) = 1$$

Las variables aleatorias X_1, X_2, \dots en la definición 5.5.1, y en las otras definiciones de esta sección, no son necesariamente independientes e idénticamente distribuidas como en una muestra aleatoria.

Frecuentemente se tiene que la secuencia de variables aleatorias corresponde a medias muestrales y que la variable aleatoria límite es constante. El resultado más famoso es el siguiente.

Teorema 5.5.1 (Ley débil de los grandes números) Sean X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas con $E[X_i] = \mu$ y $\text{Var}(X_i) = \sigma^2 < \infty$. Si se define $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ entonces, para todo $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} \Pr(|\bar{X}_n - \mu| < \epsilon) = 1$$

es decir, \bar{X}_n converge en probabilidad hacia μ .

Ejemplo 5.5.1 Suponga se tiene la secuencia X_1, X_2, \dots de variables aleatorias independientes e idénticamente distribuidas con $E[X_i] = \mu$ y $\text{Var}(X_i) = \sigma^2 < \infty$. Si se define:

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

Si se desea probar la ley débil de los grandes números para S_n^2 se usa la desigualdad de Chebychev:

$$\Pr(|S_n^2 - \sigma^2| \geq \epsilon) \leq \frac{E[(S_n^2 - \sigma^2)^2]}{\epsilon^2} = \frac{\text{Var}(S_n^2)}{\epsilon^2}$$

Una condición suficiente para que S_n^2 converja en probabilidad a σ^2 es que $\text{Var}(S_n^2) \rightarrow 0$ conforme $n \rightarrow \infty$.

Teorema 5.5.2 Si X_1, X_2, \dots converge en probabilidad hacia la variable aleatoria X y h es una función continua, entonces $h(X_1), h(X_2), \dots$ converge en probabilidad hacia $h(X)$.

5.5.2. Convergencia casi segura

Definición 5.5.2 Una secuencia de variables aleatorias, X_1, X_2, \dots converge de manera casi segura hacia la variable aleatoria X si, para todo $\epsilon > 0$

$$\Pr\left(\lim_{n \rightarrow \infty} (|X_n - X| \geq \epsilon)\right) = 1$$

Ejemplo 5.5.2 Sea el espacio muestral $S = [0, 1]$ con distribución de probabilidad uniforme. Se definen las variables aleatorias $X_n(s) = s + s^n$ y $X(s) = s$. Para todo $s \in [0, 1]$, $s^n \rightarrow 0$ conforme $n \rightarrow \infty$ y $X_n(s) \rightarrow s = X(s)$. Sin embargo, $X_n(1) = 2$ para todo n tal que $X_n(1)$ no converge a $X(1) = 1$.

Pero como la convergencia ocurre en el conjunto $[0, 1]$ y $\Pr([0, 1]) = 1$, X_n converge de forma casi segura hacia X .

Teorema 5.5.3 (Ley fuerte de los grandes números) Sean X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas con $E[X_i] = \mu$ y $\text{Var}(X_i) = \sigma^2 < \infty$ y se define $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$. Entonces, para todo $\epsilon > 0$,

$$\Pr\left(\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| < \epsilon\right) = 1$$

es decir, \bar{X}_n converge de forma casi segura hacia μ .

5.5.3. Convergencia en distribución

Definición 5.5.3 Una secuencia de variables aleatorias, X_1, X_2, \dots converge en distribución a la variable aleatoria X si:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

en todos los puntos x donde $F_X(x)$ es continua.

Ejemplo 5.5.3 Sean X_1, X_2, \dots variables aleatorias con distribución $\mathcal{U}(0, 1)$ y sea $X_{(n)} = \max_{1 \leq i \leq n} \{X_i\}$. Conforme $n \rightarrow \infty$ se espera que $X_{(n)}$ se encuentre cerca de 1, entonces para todo $\epsilon > 0$:

$$\begin{aligned} \Pr(|X_{(n)} - 1| \geq \epsilon) &= \Pr(X_{(n)} \geq 1 + \epsilon) + \Pr(X_{(n)} \leq 1 - \epsilon) \\ &= \Pr(X_{(n)} \leq 1 - \epsilon) \\ &= \Pr(X_i \leq 1 - \epsilon, i = 1, \dots, n) \\ &= (1 - \epsilon)^n \end{aligned}$$

luego $X_{(n)}$ converge en probabilidad hacia 1. Además, si se toma $\epsilon = t/n$ se tiene:

$$\Pr(X_{(n)} \leq 1 - t/n) = (1 - t/n)^n \rightarrow e^{-t}$$

lo cual es equivalente a:

$$\Pr(n(1 - X_{(n)}) \leq t) \rightarrow 1 - e^{-t}$$

es decir, la variable aleatoria $n(1 - X_{(n)})$ converge en distribución a la variable aleatoria $\mathcal{E}(1)$.

Teorema 5.5.4 (Teorema central del límite) Sea X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas cuyas funciones generatrices de momentos existen en alguna vecindad de 0 (esto es, $M_{X_i}(t)$ existe para $|t| < h$ para algún h positivo). Sea $E[X_i] = \mu$ y $\text{Var}(X_i) = \sigma^2 > 0$ (μ y σ^2 son finitos desde que la función generatriz de momentos existe). Se define $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ y sea $G_n(x)$ la función de distribución acumulada de $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Entonces para $-\infty < x < \infty$:

$$\lim_{n \rightarrow \infty} G_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

esto es, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ tiene distribución límite normal estándar.

Prueba: Se probará que para $|t| < h$, la función generatriz de momentos de $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ converge hacia $e^{t^2/2}$, la función generatriz de momentos de una variable aleatoria $N(0, 1)$. Se define $Y_i = (X_i - \mu)/\sigma$ y sea $M_Y(t)$ la función generatriz de momentos de las Y_i 's, las cuales existen para $|t| < \sigma h$. Sea W tal que:

$$W = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$$

Luego:

$$M_W(t) = M_{\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i}(t) = M_{\sum_{i=1}^n Y_i}(t/\sqrt{n}) = [M_Y(t/\sqrt{n})]^n$$

Se expande $M_Y(t/\sqrt{n})$ en una serie de potencias de Taylor alrededor de 0. Entonces:

$$M_Y(t/\sqrt{n}) = \sum_{k=0}^{\infty} M_Y^{(k)}(0) \frac{(t/\sqrt{n})^k}{k!} \text{ donde } M_Y^{(k)}(0) = \left. \frac{d^k}{dt^k} M_Y(t) \right|_{t=0}$$

Como la función generatriz de momento existe para $|t| < h$, la expresión en serie de potencias es válida si $t < \sqrt{n}\sigma h$. Usando el hecho que $M_Y^{(0)} = 1$, $M_Y^{(1)} = 0$ y $M_Y^{(2)} = 1$, ya que por construcción la media y varianza de Y son 0 y 1 respectivamente. Entonces:

$$\begin{aligned} M_Y(t/\sqrt{n}) &= M_Y^{(0)}(0) \frac{(t/\sqrt{n})^0}{0!} + \dots + M_Y^{(2)}(0) \frac{(t/\sqrt{n})^2}{2!} + R_Y(t/\sqrt{n}) \\ &= 1 + \frac{(t/\sqrt{n})^2}{2} + R_Y(t/\sqrt{n}) \end{aligned}$$

donde R_Y es el residuo en la expansión de Taylor. Una aplicación del teorema 7.4.1 de Taylor muestra que , para $t \neq 0$ fijo, se tiene:

$$\lim_{n \rightarrow \infty} \frac{R_Y(t/\sqrt{n})}{(t/\sqrt{n})^2} = 0 \text{ entonces } \lim_{n \rightarrow \infty} nR_Y(t/\sqrt{n}) = 0$$

Luego:

$$\begin{aligned} \lim_{n \rightarrow \infty} [M_Y(t/\sqrt{n})]^n &= \lim_{n \rightarrow \infty} \left[1 + \frac{(t/\sqrt{n})^2}{2} + R_Y(t/\sqrt{n}) \right]^n \\ &= \lim_{n \rightarrow \infty} \left[1 + \frac{1}{n} \left(\frac{t^2}{2} + nR_Y(t/\sqrt{n}) \right) \right]^n \end{aligned}$$

y usando el lema 2.3.1 se tiene:

$$\lim_{n \rightarrow \infty} [M_Y(t/\sqrt{n})]^n = e^{t^2/2}$$

que es la función generatriz de momentos de la distribución $\mathcal{N}(0, 1)$.

Teorema 5.5.5 (Forma fuerte del teorema central del límite) Sean X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas con $E[X_i] = \mu$ y $\text{Var}[X_i] = \sigma^2 < \infty$. Se define $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ y sea $G_n(x)$ la función de distribución acumulada de $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Entonces para $-\infty < x < \infty$:

$$\lim_{n \rightarrow \infty} G_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

esto es, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ tiene distribución límite normal estándar.

Ejemplo 5.5.4 Suponga que X_1, \dots, X_n es una muestra aleatoria de una distribución $\mathcal{BN}(r, p)$. Recordar que:

$$E[X] = \frac{r(1-p)}{p} \quad \text{y} \quad \text{Var}(X) = \frac{r(1-p)}{p^2}$$

El teorema central del límite dice que:

$$\frac{\sqrt{n}(\bar{X} - r(1-p)/p)}{\sqrt{r(1-p)/p^2}}$$

tiene distribución aproximada $\mathcal{N}(0, 1)$. Por ejemplo si $r = 10$, $p = 1/2$ y $n = 30$ entonces:

$$\begin{aligned}\Pr(\bar{X} \leq 11) &= \Pr\left(\sum_{i=1}^{30} X_i \leq 330\right) \\ &= \sum_{i=1}^{330} \binom{300 + x - 1}{x} \left(\frac{1}{2}\right)^{300} \left(\frac{1}{2}\right)^x \\ &= 0,8916\end{aligned}$$

ya que $\sum_{i=1}^n X_i \sim \mathcal{BN}(nr, p)$. Usando el teorema central de límite:

$$\begin{aligned}\Pr(\bar{X} \leq 11) &= \Pr\left(\frac{\sqrt{30}(\bar{X} - 10)}{\sqrt{20}} \leq \frac{\sqrt{30}(11 - 10)}{\sqrt{20}}\right) \\ &\cong \Pr(Z \leq 1,2247) \\ &\cong 0,8888\end{aligned}$$

Teorema 5.5.6 (Teorema de Slutsky) Si $X_n \rightarrow X$ en distribución y $Y_n \rightarrow a$ en probabilidad, donde a es constante, entonces:

- a. $Y_n X_n \rightarrow aX$ en distribución.
- b. $X_n + Y_n \rightarrow X + a$ en distribución.

Ejemplo 5.5.5 Suponga que:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \rightarrow \mathcal{N}(0, 1)$$

pero el valor de σ no se conoce. Como se vió en el ejemplo 5.5.1, si $\lim_{n \rightarrow \infty} S_n^2 = 0$, entonces $S_n^2 \rightarrow \sigma^2$ en probabilidad. Se puede probar que $\sigma/S_n \rightarrow 1$ en probabilidad y según el teorema de Slutsky:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sigma}{S_n} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \rightarrow \mathcal{N}(0, 1)$$

Capítulo 6

Principios de reducción de la data

6.1. Introducción

Un experimentador usa la información en una muestra X_1, \dots, X_n para realizar el proceso de inferencia sobre algún parámetro desconocido θ . Si el tamaño de muestra es grande entonces los valores observados en la muestra x_1, \dots, x_n podrían ser difíciles de interpretar de forma individual. Es necesario resumir la información en la muestra para determinar algunas características de interés a través del cálculo de *estadísticas* como la media, la varianza, el máximo, el mínimo, la mediana, etc.

6.2. El principio de suficiencia

Una *estadística suficiente* para un parámetro θ es una estadística que, de cierta forma, captura toda la información acerca de θ contenida en la muestra. No es posible obtener información adicional en la muestra, además del valor de la estadística suficiente. Estas consideraciones nos llevan a la técnica de reducción de datos conocida como el principio de suficiencia: “*si $T(\mathbf{X})$ es una estadística suficiente para θ , entonces el proceso de inferencia sobre θ depende de la muestra \mathbf{X} solo a través del valor $T(\mathbf{X})$* ”.

6.2.1. Estadística suficiente

Definición 6.2.1 Una estadística $T(\mathbf{X})$ es una *estadística suficiente* para θ si la distribución de la muestra \mathbf{X} dado el valor de $T(\mathbf{X})$ no depende de θ .

Teorema 6.2.1 Si $f(\mathbf{x}|\theta)$ es la función de probabilidad o densidad conjunta de \mathbf{X} , y $q(t|\theta)$ es la función de probabilidad o densidad de $T(\mathbf{X})$, entonces $T(\mathbf{X})$ es una estadística suficiente para θ si y solo si:

$$\frac{f(\mathbf{x}|\theta)}{q(T(\mathbf{x}|\theta))}$$

no depende de θ para todo \mathbf{X} .

Ejemplo 6.2.1 Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas según $\mathcal{B}(\theta)$, $0 < \theta < 1$. Para probar que $T(\mathbf{X}) = X_1 + X_2 + \dots + X_n$ es una estadística suficiente para θ hay que recordar que $T(\mathbf{X}) \sim \mathcal{BI}(n, \theta)$. Si se define $t = \sum x_i$:

$$\begin{aligned} \frac{f(\mathbf{x}|\theta)}{q(T(\mathbf{x}|\theta))} &= \frac{\prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}} \\ &= \frac{\theta^{\sum x_i} (1-\theta)^{n-\sum x_i}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}} \\ &= \frac{1}{\binom{n}{t}} \end{aligned}$$

entonces $T(\mathbf{X}) = X_1 + X_2 + \dots + X_n$ es una estadística suficiente para θ .

Ejemplo 6.2.2 Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas según $\mathcal{N}(\mu, \sigma^2)$, donde σ^2 es conocido. Se desea probar si $T(\mathbf{X}) = \bar{X}$ es una estadística suficiente para μ , entonces:

$$\begin{aligned} f(\mathbf{x}|\mu) &= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2 \right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right) \right\} \end{aligned}$$

además $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, luego:

$$\begin{aligned} \frac{f(\mathbf{x}|\theta)}{q(T(\mathbf{x}|\theta))} &= \frac{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} (\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2)\right\}}{n^{1/2}(2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right\}} \\ &= n^{-1/2}(2\pi\sigma^2)^{-(n-1)/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2\right\} \end{aligned}$$

entonces \bar{X} es una estadística suficiente para μ .

Teorema 6.2.2 (Teorema de Factorización) Sea $f(\mathbf{x}|\theta)$ la función de probabilidad o densidad conjunta de la muestra \mathbf{X} . La estadística $T(\mathbf{X})$ es suficiente para θ si y solo si existen funciones $g(t|\theta)$ y $h(\mathbf{x})$ tales que:

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}) \quad (6.2.1)$$

Ejemplo 6.2.3 Considere nuevamente el ejemplo 6.2.2:

$$\begin{aligned} f(\mathbf{x}|\mu) &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2\right\} \exp\left\{-\frac{n(\bar{x} - \mu)^2}{2\sigma^2}\right\} \\ &= h(\mathbf{x})g(T(\mathbf{x})|\mu) \end{aligned}$$

entonces \bar{X} es una estadística suficiente para μ .

Ejemplo 6.2.4 Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas según $\mathcal{U}(0, \theta)$ cuya función de densidad es:

$$f(x|\theta) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta$$

luego:

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^n \frac{1}{\theta} I_{(0,\theta)}(x_i) \\ &= \frac{1}{\theta^n} \prod_{i=1}^n I_{(0,\theta)}(x_i) \\ &= \frac{1}{\theta^n} I_{(0,\theta)}(x_{(n)}) \end{aligned}$$

entonces $T(\mathbf{X}) = X_{(n)}$ es una estadística suficiente para θ .

Ejemplo 6.2.5 Nuevamente asuma que X_1, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas según $\mathcal{N}(\mu, \sigma^2)$, donde ambos parámetros son desconocidos, es decir $\theta = (\mu, \sigma^2)$. Según el teorema de factorización:

$$\begin{aligned} f(\mathbf{x}|\mu, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right) \right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left(n(\bar{x} - \mu)^2 + (n-1)s^2 \right) \right\} \end{aligned}$$

Si $T_1(\mathbf{x}) = \bar{x}$ y $T_2(\mathbf{x}) = s^2$, entonces:

$$\begin{aligned} f(\mathbf{x}|\mu, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left(n(t_1 - \mu)^2 + (n-1)t_2 \right) \right\} \\ &= g(T_1(\mathbf{x}), T_2(\mathbf{x})|\mu, \sigma^2)h(\mathbf{x}) \end{aligned}$$

luego $T(\mathbf{x}) = (T_1(\mathbf{x}), T_2(\mathbf{x})) = (\bar{X}, S^2)$ es una estadística suficiente para el modelo normal.

Teorema 6.2.3 Sean X_1, \dots, X_n observaciones independientes e idénticamente distribuidas según la función de probabilidad o densidad $f(x|\theta)$. Suponga que $f(x|\theta)$ proviene de una familia exponencial dada por:

$$f(x|\theta) = h(x)c(\theta) \exp \left\{ \sum_{i=1}^k w_i(\theta)t_i(x) \right\}$$

donde $\theta = (\theta_1, \theta_2, \dots, \theta_d)$, $d \leq k$. Entonces:

$$T(\mathbf{X}) = \left(\sum_{j=1}^n t_1(X_j), \dots, \sum_{j=1}^n t_k(X_j) \right)$$

es una estadística suficiente para θ .

Ejemplo 6.2.6 Sean X_1, \dots, X_n observaciones independientes e idénticamente distribuidas según $\mathcal{P}(\lambda)$ cuya función de probabilidad pertenece a una familia exponencial:

$$f(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{1}{x!}e^{-\lambda} \exp \{x \log \lambda\}$$

entonces $T(\mathbf{X}) = \sum_{i=1}^n X_i$ es una estadística suficiente para λ .

6.2.2. Estadística minimal suficiente

Como es posible encontrar muchas estadísticas suficientes en un mismo problema sera necesario establecer cuál es la mejor. Recordar que el propósito de una estadística suficiente es lograr resumir la data sin pérdida de información acerca del parámetro θ , es decir que se debe buscar aquella estadística que logre la mayor reducción de data reteniendo aún toda la información sobre θ .

Definición 6.2.2 Una estadística suficiente $T(\mathbf{X})$ es llamada *estadística minimal suficiente* si, para cualquier otra estadística suficiente $T'(\mathbf{X})$, $T(\mathbf{x})$ es función de $T'(\mathbf{x})$.

Ejemplo 6.2.7 Sean X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{N}(\mu, \sigma^2)$ donde σ^2 es conocido. Se sabe que $T(\mathbf{X}) = \bar{X}$ es una estadística suficiente para μ y $T'(\mathbf{X}) = (\bar{X}, S^2)$ es una estadística suficiente para μ en el mismo problema. Claramente $T(\mathbf{X})$ logra una mayor reducción de data que $T'(\mathbf{X})$. Puede escribirse $T(\mathbf{X})$ como función de $T'(\mathbf{X})$ a través de $r(a, b) = a$ entonces:

$$T(\mathbf{x}) = \bar{x} = r(\bar{x}, s^2) = rT'(\mathbf{x})$$

Teorema 6.2.4 Sea $f(x|\theta)$ la función de probabilidad o densidad de una muestra \mathbf{X} . Suponga que existe una función $T(\mathbf{X})$ tal que, para dos puntos muestrales \mathbf{x} y \mathbf{y} , el ratio $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ no depende de θ si y solo si $T(\mathbf{x}) = T(\mathbf{y})$. Entonces $T(\mathbf{X})$ es una estadística minimal suficiente para θ .

Ejemplo 6.2.8 Sean X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{N}(\mu, \sigma^2)$ ambos parámetros desconocidos. Sean \mathbf{x} y \mathbf{y} dos muestras tales que (\bar{x}, s_X^2) y (\bar{y}, s_Y^2) son las medias y varianzas muestrales respectivamente. Entonces, usando 6.2.4:

$$\begin{aligned} \frac{f(\mathbf{x}|\mu, \sigma^2)}{f(\mathbf{y}|\mu, \sigma^2)} &= \frac{(2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (n(\bar{x} - \mu)^2 + (n-1)s_X^2) \right\}}{(2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (n(\bar{y} - \mu)^2 + (n-1)s_Y^2) \right\}} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} \left(-n(\bar{x}^2 - \bar{y}^2) + 2n\mu(\bar{x} - \bar{y}) - (n-1)(s_X^2 - s_Y^2) \right) \right\} \end{aligned}$$

El ratio anterior no dependerá de μ y σ^2 si y solo si $\bar{x} = \bar{y}$ y $s_X^2 = s_Y^2$. Luego por el teorema 6.2.4 (\bar{X}, S^2) es una estadística minimal suficiente para (μ, σ^2) .

Ejemplo 6.2.9 Suponga X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{U}(\theta, \theta + 1)$, $-\infty < \theta < \infty$ entonces la función de densidad conjunta de \mathbf{X} es:

$$\prod_{i=1}^n I_{(\theta, \theta+1)}(x_i) = I_{(x_{(n)}-1, x_{(1)})}(\theta)$$

Luego:

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \frac{I_{(x_{(n)}-1, x_{(1)})}(\theta)}{I_{(y_{(n)}-1, y_{(1)})}(\theta)}$$

no depende de θ si $x_{(n)} = y_{(n)}$ y $x_{(1)} = y_{(1)}$. Luego $T(\mathbf{X}) = (X_{(1)}, X_{(n)})$ es una estadística minimal suficiente.

Una estadística minimal suficiente no es única. Cualquier función uno a uno de una estadística minimal suficiente es también una estadística minimal suficiente. Luego, $T'(\mathbf{X}) = (X_{(n)} - X_{(1)}, (X_{(1)} + X_{(n)})/2)$ es también una estadística minimal suficiente en el ejemplo 6.2.9 y $T'(\mathbf{X}) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ es también una estadística minimal suficiente en el ejemplo 6.2.8.

6.2.3. Estadística ancillar

En las secciones anteriores se consideraron las estadísticas suficientes que contienen toda la información sobre θ en la muestra. En esta sección se introduce un tipo diferente de estadística que tiene un propósito complementario.

Definición 6.2.3 Una estadística $S(\mathbf{X})$ cuya distribución no depende del parámetro θ es llamada *estadística ancillar*.

Ejemplo 6.2.10 Sean X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{U}(\theta, \theta + 1)$. Sean $X_{(1)} < \dots < X_{(n)}$ las estadísticas de orden de la muestra. Si $R = X_{(n)} - X_{(1)}$ se tiene que:

$$F(x|\theta) = \begin{cases} 0 & x \leq \theta \\ x - \theta & \theta < x < \theta + 1 \\ 1 & x \geq \theta + 1 \end{cases}$$

La función de densidad conjunta de $X_{(1)}$ y $X_{(n)}$ es:

$$f_{X_{(1)}, X_{(n)}}(u, v|\theta) = \begin{cases} n(n-1)(v-u)^{n-2} & \theta < u < v < \theta + 1 \\ 0 & \text{de otro modo} \end{cases}$$

Usando las transformaciones $R = X_{(n)} - X_{(1)}$ y $M = (X_{(1)} + X_{(n)})/2$ se tiene $X_{(1)} = (2M - R)/2$ y $X_{(n)} = (2M + R)/2$ tal que $J = 1$. Luego:

$$f_{R,M}(r, m|\theta) = \begin{cases} n(n-1)r^{n-2} & \theta + (r/2) < m < \theta + 1 - (r/2) \\ 0 & \text{de otro modo} \end{cases}$$

La distribución marginal:

$$f_R(r|\theta) = \int_{\theta+(r/2)}^{\theta+1-(r/2)} n(n-1)r^{n-2} dm = n(n-1)r^{n-2}(1-r), \quad 0 < r < 1$$

Es decir, R tiene distribución $\mathcal{BE}(\alpha = n-1, \beta = 2)$ y no depende de θ por lo que R es una estadística ancillar.

6.2.4. Estadística suficiente, ancillar y completa

Definición 6.2.4 Sea $f(t|\theta)$ una familia con función de probabilidad o densidad para una estadística $T(\mathbf{X})$. La familia de distribución de probabilidad es llamada *completa* si $E_\theta[g(T)] = 0$ para todo θ implica que $\Pr_\theta[g(T) = 0] = 1$ para todo θ . Equivalentemente, $T(\mathbf{X})$ es llamada una *estadística completa*.

Ejemplo 6.2.11 Sea $T \sim \mathcal{BI}(n, p)$, $0 < p < 1$. Sea g una función tal que $E_p[g(T)] = 0$, es decir:

$$E_p[g(T)] = \sum_{t=0}^n g(t) \binom{n}{t} p^t (1-p)^{n-t} = (1-p)^n \sum_{t=0}^n g(t) \binom{n}{t} \left(\frac{p}{1-p}\right)^t$$

luego $\sum_{t=0}^n g(t) \binom{n}{t} \left(\frac{p}{1-p}\right)^t = 0$ entonces $g(t) = 0$ para $t = 0, 1, 2, \dots, n$ y $\Pr_p(g(T) = 0) = 1$, para todo p . Luego T es una estadística completa.

Ejemplo 6.2.12 Sean X_1, \dots, X_n observaciones independientes e idénticamente distribuidas como $\mathcal{U}(0, \theta)$, $0 < \theta < \infty$. Se tiene que $T(\mathbf{X}) = X_{(n)}$ es una estadística suficiente y que su función de densidad es:

$$f(t|\theta) = \begin{cases} nt^{n-1}\theta & 0 < t < \theta \\ 0 & \text{de otro modo} \end{cases}$$

Suponga $g(t)$ es una función que satisface $E_\theta [g(T)] = 0$ para todo θ . Como $E_\theta [g(T)] = 0$ es una función constante su derivada con respecto a θ es cero, entonces se tiene:

$$0 = \frac{d}{d\theta} E_\theta [g(T)] = \frac{d}{d\theta} \int_0^\theta g(t) n t^{n-1} \theta^{-n} dt$$

Para toda función Riemman integrable $\frac{d}{d\theta} \int_0^\theta g(t) dt = g(\theta)$. Luego:

$$\begin{aligned} &= \frac{d}{d\theta} \theta^{-n} \int_0^\theta g(t) n t^{n-1} dt \\ &= \theta^{-n} \frac{d}{d\theta} \int_0^\theta g(t) n t^{n-1} dt + \left(\frac{d}{d\theta} \theta^{-n} \right) \int_0^\theta n g(t) t^{n-1} dt \\ &= (\theta^{-n}) g(\theta) n \theta^{n-1} + 0 \\ &= \frac{1}{\theta} g(\theta) n = 0 \end{aligned}$$

entonces $g(\theta) = 0$, luego T es una estadística completa.

Teorema 6.2.5 Sean X_1, \dots, X_n observaciones independientes e idénticamente distribuidas de una familia exponencial con función de probabilidad o densidad o función de probabilidad de la forma:

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp \left\{ \sum_{j=1}^k w_j(\boldsymbol{\theta}) t_j(x) \right\}$$

donde $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$, entonces la estadística

$$T(\mathbf{X}) = \left(\sum_{i=1}^n t_1(X_i), \dots, \sum_{i=1}^n t_k(X_i) \right)$$

es completa si $\{(w_1(\boldsymbol{\theta}), \dots, w_k(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta)\}$ contiene un conjunto abierto en \Re^k .

Ejemplo 6.2.13 En el ejemplo 6.2.6 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ es una estadística suficiente y completa para λ .

Capítulo 7

Estimación puntual

7.1. Introducción

Definición 7.1.1 Un *estimador puntual* es cualquier función $W(X_1, \dots, X_n)$ de la muestra. Es decir, cualquier estadística es un estimador puntual.

Se debe tener clara la diferencia entre *estimador* y *estimación*. Un *estimador* es una función de una muestra, mientras que una *estimación* es el valor obtenido al aplicar un *estimador* a los datos de una muestra. Es decir, un estimador es una función de las variables aleatorias X_1, \dots, X_n mientras que una estimación es una función de los valores muestrales x_1, \dots, x_n .

7.2. Métodos para encontrar estimadores

En muchos casos habrá un candidato evidente o natural para ser el estimador puntual de un parámetro particular y a menudo la intuición puede inducirnos a obtener buenos estimadores. Por ejemplo, la media muestral es un candidato natural para estimar la media poblacional.

7.2.1. Métodos de momentos

Sea X_1, \dots, X_n una muestra de una población con función de probabilidad o densidad $f(x|\theta_1, \dots, \theta_k)$. Los estadísticos por el método de momentos se encuentran igualando los k primeros momentos muestrales a sus correspondientes k momentos poblacionales y resolviendo simultáneamente las

ecuaciones. Es decir, se definen:

$$\begin{aligned} m_1 &= \frac{1}{n} \sum_{i=1}^n X_i^1, & \mu_1 &= E[X^1] \\ m_2 &= \frac{1}{n} \sum_{i=1}^n X_i^2, & \mu_2 &= E[X^2] \\ & & \vdots & \\ m_k &= \frac{1}{n} \sum_{i=1}^n X_i^k, & \mu_k &= E[X^k] \end{aligned}$$

El momento poblacional μ_j es, por lo general, una función de $\theta_1, \dots, \theta_k$ digamos $\mu_j(\theta_1, \dots, \theta_k)$. El estimador por el método de momentos $(\tilde{\theta}_1, \dots, \tilde{\theta}_k)$ de $(\theta_1, \dots, \theta_k)$ se obtiene resolviendo el siguiente sistema de ecuaciones en términos de (m_1, \dots, m_k) :

$$\begin{aligned} m_1 &= \mu_1(\theta_1, \dots, \theta_k) \\ m_2 &= \mu_2(\theta_1, \dots, \theta_k) \\ &\vdots \\ m_k &= \mu_k(\theta_1, \dots, \theta_k) \end{aligned}$$

Ejemplo 7.2.1 Suponga X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{N}(\mu, \sigma^2)$. Se tiene $m_1 = \bar{X} = \mu_1$, $m_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 = \mu_2$ y resolviendo:

$$\tilde{\mu} = \bar{X} \quad \text{y} \quad \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

En este caso, el estimador por el método de momentos coincide con nuestra intuición sobre el candidato natural para μ .

Ejemplo 7.2.2 Suponga X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{BI}(k, p)$, ambos parámetros desconocidos:

$$\begin{aligned} m_1 &= \frac{1}{n} \sum_{i=1}^n X_i = kp = E[X] \\ m_2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 = kp(1-p) + k^2p^2 = E[X^2] \end{aligned}$$

resolviendo :

$$\tilde{k} = \frac{\overline{X}^2}{\overline{X} - \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2} \quad \text{y} \quad \tilde{p} = \frac{\overline{X}}{\tilde{k}}$$

7.2.2. Estimadores de máxima verosimilitud

Definición 7.2.1 Sea $f(\mathbf{x}|\theta)$ que denota la función de probabilidad o densidad conjunta de la muestra $\mathbf{X} = (X_1, \dots, X_n)$. Entonces, dado que $\mathbf{X} = \mathbf{x}$ es observado, la función de θ definida por:

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

es llamada *función de verosimilitud*.

Si X_1, \dots, X_n es una muestra independiente e idénticamente distribuida de una población con función de probabilidad o densidad $f(\mathbf{x}|\theta_1, \dots, \theta_k)$, la *función de verosimilitud* se define por:

$$L(\theta|\mathbf{x}) = L(\theta_1, \dots, \theta_k|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta_1, \dots, \theta_k) \quad (7.2.1)$$

Definición 7.2.2 Para cada punto muestral \mathbf{x} , sea $\hat{\theta}(\mathbf{x})$ el valor del parámetro en que $L(\theta|\mathbf{x})$ toma su máximo valor como función de θ , con \mathbf{x} fijo. Un *estimador de máxima verosimilitud* del parámetro θ basado en la muestra \mathbf{X} es $\hat{\theta}(\mathbf{X})$.

Si la función de verosimilitud es diferenciable en θ_i , los posibles candidatos para estimadores de máxima verosimilitud son los valores de $(\theta_1, \dots, \theta_k)$ que resuelven:

$$\frac{\partial}{\partial \theta_i} L(\theta|\mathbf{x}) = 0, \quad i = 1, \dots, k \quad (7.2.2)$$

Ejemplo 7.2.3 Suponga X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{N}(\theta, \sigma^2 = 1)$ entonces:

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}(x_i - \theta)^2} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2}$$

Como $\frac{d}{d\theta} L(\theta|\mathbf{x}) = 0$ se reduce a $\sum_{i=1}^n (x_i - \theta) = 0$ entonces $\hat{\theta} = \bar{x}$ es un posible candidato. Luego verificando que:

$$\left. \frac{d^2}{d\theta^2} L(\theta|\mathbf{x}) \right|_{\theta=\bar{x}} < 0$$

se concluye que $\hat{\theta} = \bar{X}$ es el estimador de máxima verosimilitud.

En muchos casos es fácil trabajar con el logaritmo natural de $L(\theta|\mathbf{x})$ conocido como *logaritmo de la verosimilitud*. Lo anterior es posible debido a que la función log-verosimilitud es estrictamente decreciente sobre $(0, \infty)$.

Ejemplo 7.2.4 Sea X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{B}(p)$:

$$L(p|\mathbf{x}) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}$$

El logaritmo de la verosimilitud es:

$$\log L(p|\mathbf{x}) = \sum_{i=1}^n x_i \log p + (n - \sum_{i=1}^n x_i) \log(1-p)$$

Luego $\frac{d}{dp} \log L(p|\mathbf{x}) = 0$, entonces $\hat{p} = \bar{x}$ es el candidato a estimador de máxima verosimilitud. Luego, verificando que:

$$\left. \frac{d^2}{dp^2} \log L(p|\mathbf{x}) \right|_{p=\bar{x}} < 0$$

se concluye que $\hat{p} = \bar{X}$ es el estimador de máxima verosimilitud.

Teorema 7.2.1 (Propiedad de invariancia) Si $\hat{\theta}$ es el estimador de máxima verosimilitud de θ , entonces para toda función $\tau(\theta)$, su estimador de máxima verosimilitud es $\tau(\hat{\theta})$.

Usando el teorema anterior, se puede establecer que en el problema 7.2.3 el estimador de máxima verosimilitud de θ^2 es \bar{X}^2 . Además, el estimador de máxima verosimilitud de $\sqrt{p(1-p)}$ es $\sqrt{\bar{X}(1-\bar{X})}$ en el problema 7.2.4.

Ejemplo 7.2.5 Suponga X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{N}(\theta, \sigma^2)$, ambos parámetros desconocidos. Entonces:

$$L(\theta, \sigma^2|\mathbf{x}) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2}$$

$$\log L(\theta, \sigma^2 | \mathbf{x}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2$$

las derivadas parciales con respecto a θ y σ^2 son:

$$\frac{\partial}{\partial \theta} \log L = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \theta) = 0 \quad \text{y} \quad \frac{\partial}{\partial \sigma^2} \log L = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \theta)^2 = 0$$

entonces $\hat{\theta} = \bar{x}$ y $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$, pero ¿se trata de un máximo global? Para eso al menos una derivada parcial de segundo orden deber ser negativa:

$$\frac{\partial^2}{\partial \theta^2} \log L \Big|_{\theta=\bar{x}} = \frac{1}{\sigma^2} \sum_{i=1}^n (-1) = -\frac{n}{\sigma^2} < 0$$

y además el Jacobiano debe ser positivo:

$$\begin{vmatrix} \frac{\partial^2}{\partial \theta^2} \log L & \frac{\partial}{\partial \theta \partial \sigma^2} \log L \\ \frac{\partial}{\partial \theta \partial \sigma^2} \log L & \frac{\partial^2}{\partial (\sigma^2)^2} \log L \end{vmatrix}$$

Luego:

$$\begin{aligned} &= \begin{vmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \theta)^2 \\ -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \theta)^2 & \frac{n}{\sigma^4} - \frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \theta)^2 \end{vmatrix}_{\theta=\hat{\theta}, \sigma^2=\hat{\sigma}^2} \\ &= \frac{1}{\sigma^6} \frac{n^2}{2} \Big|_{\theta=\hat{\theta}, \sigma^2=\hat{\sigma}^2} \\ &= \frac{1}{\hat{\sigma}^6} \frac{n^2}{2} > 0 \end{aligned}$$

Finalmente, $\hat{\theta} = \bar{X}$ y $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ son los estimadores de máxima verosimilitud.

7.3. Métodos de evaluación de estimadores

7.3.1. Error cuadrático medio

Definición 7.3.1 El error cuadrático medio del estimador W de un parámetro θ se define por $E_{\theta}[(W - \theta)^2]$. Se puede probar que:

$$E_{\theta}[(W - \theta)^2] = \text{Var}_{\theta}(W) + \text{Sesgo}_{\theta}^2[W] \quad (7.3.1)$$

Definición 7.3.2 El *sesgo* del estimador puntual W del parámetro θ , es la diferencia entre su valor esperado y θ . Es decir, $\text{Sesgo}_{\theta}[W] = E_{\theta}[W] - \theta$. Un estimador cuyo sesgo es cero es llamado *insesgado*. Para estimadores insesgados se tiene $E_{\theta}[(W - \theta)^2] = \text{Var}_{\theta}(W)$.

Ejemplo 7.3.1 Suponga X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{N}(\mu, \sigma^2)$. Las estadísticas \bar{X} y S^2 son ambos estimadores insesgados ya que:

$$E[\bar{X}] = \mu \quad \text{y} \quad E[S^2] = \sigma^2$$

Lo anterior es cierto aún sin el supuesto de normalidad. El error cuadrático medio de estos estadísticos es:

$$E[(\bar{X} - \mu)^2] = \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$E[(S^2 - \sigma^2)^2] = \text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

un estimador alternativo para σ^2 es el estimador de máxima verosimilitud $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{(n-1)}{n} S^2$, entonces:

$$E[\hat{\sigma}^2] = E\left[\frac{n-1}{n} S^2\right] = \frac{n-1}{n} \sigma^2$$

es decir, $\hat{\sigma}^2$ es un estadístico sesgado de σ^2 . La variancia de $\hat{\sigma}^2$ puede calcularse como:

$$\text{Var}(\hat{\sigma}^2) = \text{Var}\left(\frac{n-1}{n} S^2\right) = \left(\frac{n-1}{n}\right)^2 \text{Var}(S^2) = \frac{2(n-1)}{n^2} \sigma^4$$

Luego:

$$\begin{aligned} \text{ECM}[\hat{\sigma}^2] &= E[(\hat{\sigma}^2 - \sigma^2)^2] \\ &= \text{Var}(\hat{\sigma}^2) + \text{Sesgo}^2[\hat{\sigma}^2] \\ &= \frac{2(n-1)}{n^2} \sigma^4 + \left(\frac{n-1}{n} \sigma^2 - \sigma^2\right)^2 \\ &= \left(\frac{2n-1}{n^2}\right) \sigma^4 \end{aligned}$$

Finalmente:

$$\text{ECM}[\hat{\sigma}^2] = \left(\frac{2n-1}{n^2}\right) \sigma^4 < \left(\frac{2}{n-1}\right) \sigma^4 = \text{ECM}[S^2]$$

7.3.2. Mejores estimadores insesgados

Definición 7.3.3 Un estimador W^* es el *mejor estimador insesgado* de $\tau(\theta)$ si satisface que $E_\theta[W^*] = \tau(\theta)$, para todo θ y para cualquier otro estimador W con $E_\theta[W] = \tau(\theta)$ se tiene $\text{Var}_\theta(W^*) \leq \text{Var}_\theta(W)$ para todo θ . W^* también es llamado *estimador insesgado uniforme de mínima variancia* de $\tau(\theta)$.

Ejemplo 7.3.2 Sean X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{P}(\lambda)$ y sean \bar{X} y S^2 . Recordar que para la función de probabilidad de Poisson, la media y la variancia son iguales a λ . Luego aplicando el teorema 5.2.2 se tiene:

$$E_\lambda[\bar{X}] = \lambda$$

$$E_\lambda[S^2] = \lambda$$

es decir que ambos son estimadores insesgados de λ . Para determinar cuál es mejor se comparan las varianzas, aplicando nuevamente el teorema 5.2.2, obteniéndose que $\text{Var}_\lambda(\bar{X}) \leq \text{Var}_\lambda(S^2)$. Aún considerando que \bar{X} es mejor que S^2 , considere la siguiente clase de estimadores:

$$W_a(\bar{X}, S^2) = a\bar{X} + (1-a)S^2$$

para toda constante a , $E_\lambda[W_a(\bar{X}, S^2)] = \lambda$, es decir se tienen infinitos estimadores insesgados de λ . La pregunta es, aún siendo \bar{X} mejor estimador que S^2 , ¿ \bar{X} es mejor que $W_a(\bar{X}, S^2)$ para todo a ?

Teorema 7.3.1 (Cramér-Rao) Sea X_1, \dots, X_n una muestra con función de probabilidad o densidad $f(\mathbf{x}|\theta)$ y sea $W(\mathbf{X}) = W(X_1, \dots, X_n)$ algún estimador donde $E_\theta[W(\mathbf{X})]$ es una función diferenciable de θ . Suponga que la función de densidad conjunta $f(\mathbf{x}|\theta) = f(x_1, \dots, x_n|\theta)$ satisface:

$$\frac{d}{d\theta} \int \cdots \int h(\mathbf{x}) f(\mathbf{x}|\theta) dx_1 \cdots dx_n = \int \cdots \int h(\mathbf{x}) \frac{\partial}{\partial \theta} f(\mathbf{x}|\theta) dx_1 \cdots dx_n \quad (7.3.2)$$

para cualquier función $h(\mathbf{x})$ con $E_\theta[|h(\mathbf{X})|] < \infty$. Entonces:

$$\text{Var}_\theta(W(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta} E_\theta[W(\mathbf{X})]\right)^2}{E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(\mathbf{X}|\theta) \right)^2 \right]} \quad (7.3.3)$$

Corolario 7.3.1 (Crámer-Rao caso independiente e idénticamente distribuido) Sean X_1, \dots, X_n independientes e idénticamente distribuidos con función de probabilidad o densidad $f(x|\theta)$ y sea $W(\mathbf{X}) = W(X_1, \dots, X_n)$ cualquier estadística tal que $E_\theta[W(\mathbf{X})]$ es una función diferenciable de θ . Si la función de densidad conjunta $f(\mathbf{x}|\theta) = \prod f(x_i|\theta)$ satisface 7.3.2 :

$$\text{Var}_\theta(W(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta} E_\theta[W(\mathbf{X})]\right)^2}{n E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X|\theta) \right)^2 \right]}$$

La cantidad $E_\theta \left[\left(\frac{\partial}{\partial \theta} \log \prod_{i=1}^n f(X_i|\theta) \right)^2 \right]$ es llamada *número de información o información de Fisher* de la muestra.

Lema 7.3.1 Si $f(x|\theta)$ satisface:

$$\frac{d}{d\theta} E_\theta \left[\frac{\partial}{\partial \theta} \log f(X|\theta) \right] = \int \frac{\partial}{\partial \theta} \left[\left(\frac{\partial}{\partial \theta} \log f(x|\theta) f(x|\theta) \right) \right] dx$$

lo cual es verdadero para una familia exponencial, entonces:

$$E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X|\theta) \right)^2 \right] = -E_\theta \left[\frac{\partial^2}{\partial \theta^2} \log f(X|\theta) \right]$$

Ejemplo 7.3.3 Volviendo al ejercicio de la distribución de Poisson:

$$\text{Var}_\lambda(W(\mathbf{X})) \geq \frac{\left(\frac{d}{d\lambda} E_\lambda[W(\mathbf{X})]\right)^2}{n E_\lambda \left[\left(\frac{\partial}{\partial \lambda} \log f(X|\lambda) \right)^2 \right]}$$

Si consideramos cualquier estimador insesgado:

$$\begin{aligned} \text{Var}_\lambda(W(\mathbf{X})) &\geq \frac{\left(\frac{d}{d\lambda}(\lambda)\right)^2}{n E_\lambda \left[\left(\frac{\partial}{\partial \lambda} \log f(X|\lambda) \right)^2 \right]} \\ &\geq \frac{1}{n E_\lambda \left[\left(\frac{\partial}{\partial \lambda} \log f(X|\lambda) \right)^2 \right]} \end{aligned}$$

y como la distribución de Poisson pertenece a una familia exponencial:

$$\begin{aligned}\text{Var}_\lambda(W(\mathbf{X})) &\geq \frac{1}{-n\text{E}_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log f(X|\lambda)\right]} \\ &\geq \frac{1}{-n(-\frac{1}{\lambda})} \\ &\geq \frac{\lambda}{n}\end{aligned}$$

Como $\text{Var}_\lambda(\bar{X}) = \lambda/n$, entonces \bar{X} es el mejor estimador insesgado de λ .

Es importante recordar que un supuesto clave en el teorema de Crámer-Rao es la posibilidad de derivar bajo el signo de la integral, algo que es en cierto modo restrictivo pero que se satisface si la densidad pertenece a una familia exponencial.

Ejemplo 7.3.4 Sean X_1, \dots, X_n independientes e idénticamente distribuidos según $\mathcal{N}(\mu, \sigma^2)$. Considere el problema de estimar σ^2 cuando μ es conocido. La función de densidad cumple con las condiciones del teorema de Crámer-Rao y el lema 7.3.2, luego:

$$\begin{aligned}\frac{\partial^2}{\partial(\sigma^2)^2}\log f &= \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6} \\ -\text{E}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\log f\right] &= -\text{E}\left[\frac{1}{2\sigma^4} - \frac{(X-\mu)^2}{\sigma^6}\right] = \frac{1}{2\sigma^4}\end{aligned}$$

Para todo estimador insesgado W de σ^2 se tiene:

$$\text{Var}(W|\mu, \sigma^2) \geq \frac{2\sigma^4}{n}$$

En el ejemplo 7.3.1 se vió que:

$$\text{Var}(S^2|\mu, \sigma^2) \geq \frac{2\sigma^4}{n-1}$$

es decir que S^2 no alcanza la cota inferior de Crámer-Rao. La pregunta ahora es, ¿existe algún estimador insesgado de σ^2 cuya varianza alcance la cota mencionada?

Corolario 7.3.2 Sean X_1, \dots, X_n independientes e idénticamente distribuidos como $f(x|\theta)$ tal que satisface las condiciones del teorema Crámer-Rao. Sea $L(\theta|x) = \prod_{i=1}^n f(x_i|\theta)$ denota la función de verosimilitud. Si $W(\mathbf{X}) = W(X_1, \dots, X_n)$ es cualquier estimador insesgado de $\tau(\theta)$, entonces $W(\mathbf{X})$ alcanza la cota inferior real de Crámer-Rao sí y solo si:

$$a(\theta)[W(\mathbf{x}) - \tau(\theta)] = \frac{\partial}{\partial \theta} \log L(\theta|\mathbf{x}) \quad (7.3.4)$$

para alguna función $a(\theta)$.

Ejemplo 7.3.5 Retomando el ejemplo 7.3.4 se tiene:

$$L(\mu, \sigma^2|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

y además:

$$\frac{\partial}{\partial \sigma^2} \log L(\mu, \sigma^2|\mathbf{x}) = \frac{n}{2\sigma^4} \left(\sum_{i=1}^n \frac{(x_i - \mu)^2}{n} - \sigma^2 \right)$$

Luego:

$$a(\sigma^2) \left(\sum_{i=1}^n \frac{(x_i - \mu)^2}{n} - \sigma^2 \right) = \frac{n}{2\sigma^4} \left(\sum_{i=1}^n \frac{(x_i - \mu)^2}{n} - \sigma^2 \right)$$

entonces:

$$a(\sigma^2) = \frac{n}{2\sigma^4}$$

Es decir, el mejor estimador insesgado de σ^2 es $\sum (x_i - \mu)^2 / n$ y puede ser calculado solo si μ es conocido, en caso contrario no puede alcanzarse la cota.

7.3.3. Suficiencia e insesgabilidad

Recordar que si X y Y son dos variables aleatorias cualesquiera entonces, siempre que los esperados existan, se tiene:

$$\begin{aligned} E[X] &= E[E[X|Y]] \\ \text{Var}[X] &= \text{Var}(E[X|Y]) + E[\text{Var}(X|Y)] \end{aligned}$$

Teorema 7.3.2 (Rao-Blackwell) Sea W un estimador insesgado de $\tau(\theta)$ y sea T una estadística suficiente para θ . Se define $\phi(T) = E[W|T]$. Entonces $E_\theta[\phi(T)] = \tau(\theta)$ y $\text{Var}_\theta(\phi(T)) \leq \text{Var}_\theta(W)$ para todo θ , es decir, $\phi(T)$ es el mejor estimador insesgado uniforme de $\tau(\theta)$.

Teorema 7.3.3 Si W es el mejor estimador insesgado de $\tau(\theta)$, entonces W es único.

Teorema 7.3.4 Sea T una estadística suficiente y además completa para un parámetro θ y sea $\phi(T)$ algún estimador basado sólo en T . Entonces $\phi(T)$ es el único mejor estimador insesgado de su valor esperado.

Ejemplo 7.3.6 Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas según $\mathcal{BI}(k, \theta)$. Se desea estimar la probabilidad de obtener un éxito, es decir:

$$\tau(\theta) = \Pr_\theta(X = 1) = k\theta(1 - \theta)^{k-1}$$

Además $\sum_{i=1}^n X_i \sim \mathcal{BI}(kn, \theta)$ es una estadística suficiente y completa. El siguiente estimador:

$$h(X_1) = \begin{cases} 1 & \text{si } X_1 = 1 \\ 0 & \text{de otro modo} \end{cases}$$

satisface $E_\theta[h(X_1)] = k\theta(1 - \theta)^{k-1}$ por lo que se trata de un estimador insesgado de $\tau(\theta)$. Luego:

$$\phi\left(\sum_{i=1}^n X_i\right) = E\left[h(X_1) \mid \sum_{i=1}^n X_i\right]$$

es el mejor estimador insesgado de $\tau(\theta)$. Suponga que se observa $\sum_{i=1}^n X_i = t$, entonces:

$$\phi(t) = E\left[h(X_1) \mid \sum_{i=1}^n X_i = t\right] = \Pr_\theta(X_1 = 1 \mid \sum_{i=1}^n X_i = t)$$

aplicando la definición de probabilidad condicional:

$$\begin{aligned}
\phi(t) &= \frac{\Pr_\theta(X_1 = 1, \sum_{i=1}^n X_i = t)}{\Pr_\theta(\sum_{i=1}^n X_i = t)} \\
&= \frac{\Pr_\theta(X_1 = 1, \sum_{i=2}^n X_i = t-1)}{\Pr_\theta(\sum_{i=1}^n X_i = t)} \\
&= \frac{\Pr_\theta(X_1 = 1) \Pr_\theta(\sum_{i=2}^n X_i = t-1)}{\Pr_\theta(\sum_{i=1}^n X_i = t)}
\end{aligned}$$

Además $X_i \sim \mathcal{BI}(k, \theta)$ y $\sum_{i=2}^n X_i \sim \mathcal{BI}(k(n-1), \theta)$. Entonces:

$$\begin{aligned}
\phi(t) &= \frac{\left[k\theta(1-\theta)^{k-1} \right] \binom{k(n-1)}{t-1} \theta^{t-1} (1-\theta)^{k(n-1)-(t-1)}}{\binom{kn}{t} \theta^t (1-\theta)^{kn-t}} \\
&= k \frac{\binom{k(n-1)}{t-1}}{\binom{kn}{t}}
\end{aligned}$$

es el mejor estimador insesgado de $\tau(\theta) = \Pr(X = 1)$.

7.3.4. Consistencia

Definición 7.3.4 Una secuencia de estimadores $W_n = W_n(X_1, \dots, X_n)$ es una *secuencia consistente de estimadores* del parámetro θ si, para todo $\epsilon > 0$ y todo $\theta \in \Theta$:

$$\lim_{n \rightarrow \infty} \Pr_\theta(|W_n - \theta| < \epsilon) = 1 \quad (7.3.5)$$

Ejemplo 7.3.7 Sean X_1, \dots, X_n independientes e idénticamente distribuidas según $\mathcal{N}(\theta, 1)$ y considere la secuencia:

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

recodando que $\bar{X}_n \sim \mathcal{N}(\theta, \frac{1}{n})$ se tiene:

$$\begin{aligned}
 \Pr_{\theta}(|\bar{X}_n - \theta| < \epsilon) &= \int_{\bar{x}_n = \theta - \epsilon}^{\bar{x}_n = \theta + \epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{n}{2}(\bar{x}_n - \theta)^2} d\bar{x}_n \\
 &= \int_{y = -\epsilon}^{y = \epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{n}{2}y^2} dy \quad (y = \bar{x}_n - \theta) \\
 &= \int_{t = -\epsilon\sqrt{n}}^{t = \epsilon\sqrt{n}} \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}t^2} dt \quad (t = y\sqrt{n}) \\
 &= \Pr_{\theta}(-\epsilon\sqrt{n} < Z < \epsilon\sqrt{n}) \rightarrow 1
 \end{aligned}$$

cuando $n \rightarrow \infty$. Luego \bar{X}_n es una secuencia consistente de estimadores de θ .

Teorema 7.3.5 Si W_n es una secuencia de estimadores del parámetro θ que satisfacen

- a. $\lim_{n \rightarrow \infty} \text{Var}_{\theta}[W_n] = 0$.
- b. $\lim_{n \rightarrow \infty} \text{Sesgo}_{\theta}[W_n] = 0$.

entonces W_n es una secuencia consistente de estimadores de θ .

Ejemplo 7.3.8 Como:

$$E_{\theta}[\bar{X}_n] = \theta \text{ y } \text{Var}_{\theta}(\bar{X}_n) = \frac{1}{n}$$

satisfacen las condiciones del teorema 7.3.5, luego la secuencia \bar{X}_n es consistente. Además, del teorema 5.2.2, \bar{X}_n proviene de un muestreo independiente e idénticamente distribuido de cualquier población con media θ por lo que es consistente para dicho parámetro siempre que la varianza sea finita.

Teorema 7.3.6 Si W_n una secuencia consistente de estimadores del parámetro θ . Sean a_1, a_2, \dots y b_1, b_2, \dots secuencias de constantes que satisfacen:

- a. $\lim_{n \rightarrow \infty} a_n = 1$.

b. $\lim_{n \rightarrow \infty} b_n = 0$.

entonces, $U_n = a_n W_n + b_n$ es una secuencia consistente de estimadores de θ .

Teorema 7.3.7 (Consistencia de los estimadores de máxima verosimilitud) Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas de $f(x|\theta)$, y sea $L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$ la función de verosimilitud. Sea $\hat{\theta}$ el estimador de máxima verosimilitud de θ y $\tau(\theta)$ una función continua de θ . Bajo ciertas condiciones de regularidad sobre $f(x|\theta)$, y por consiguiente $L(\theta|\mathbf{x})$, para todo $\epsilon > 0$ y $\theta \in \Theta$,

$$\lim_{n \rightarrow \infty} \Pr_{\theta} (|\tau(\hat{\theta}) - \tau(\theta)| \geq \epsilon) = 0$$

es decir, $\tau(\hat{\theta})$ es un estimador consistente de $\tau(\theta)$.

7.4. Otras consideraciones

7.4.1. Variancia asintótica de los estimadores de máxima verosimilitud

Definición 7.4.1 Una secuencia de estimadores W_n es *asintóticamente eficiente* para un parámetro $\tau(\theta)$ si:

$$\lim_{n \rightarrow \infty} \frac{\text{Var}_{\theta}(W_n)}{\left[\frac{[\tau'(\theta)]^2}{n \mathbb{E}_{\theta} \left[\left(\frac{\partial}{\partial \theta} \log f(x|\theta) \right)^2 \right]} \right]} = 1$$

esto es, W_n alcanza la cota inferior de Crámer-Rao conforme $n \rightarrow \infty$.

7.4.2. Aproximación por series de Taylor

Definición 7.4.2 Si una función $g(x)$ tiene derivadas de orden r , es decir que existe $g^{(r)}(x) = \frac{d^r}{dx^r} g(x)$, entonces para cualquier constante a , la *polinomial de Taylor de orden r alrededor de a* es:

$$T_r(x) = \sum_{i=0}^r \frac{g^{(i)}(a)}{i!} (x - a)^i$$

Teorema 7.4.1 (Taylor) Si $g^{(r)}(a) = \frac{d^r}{dx^r}g(x) \mid_{x=a}$ existe, entonces:

$$\lim_{x \rightarrow a} \frac{g(x) - T_r(x)}{(x - a)^r} = 0$$

Capítulo 8

Prueba de hipótesis

8.1. Introducción

Definición 8.1.1 Una hipótesis es una afirmación acerca de un parámetro.

Definición 8.1.2 Las dos hipótesis complementarias en un problema de prueba de hipótesis son llamadas *hipótesis nula* e *hipótesis alternativa* y se denotan por H_0 y H_1 respectivamente.

El objetivo de una prueba de hipótesis es decidir, en base a una muestra extraída a partir de la población, cuál de las dos hipótesis complementarias es verdadera.

Definición 8.1.3 Una prueba de hipótesis es una regla que especifica:

- a. Para que valores muestrales la decisión es no rechazar H_0 y suponer que es verdadera.
- b. Para que valores muestrales la decisión es rechazar H_0 y aceptar H_1 como verdadera.

El subconjunto del espacio muestral para el que se rechaza H_0 se denomina *región de rechazo* o *región crítica*. El complemento de la región de rechazo es llamado la región de no rechazo.

8.2. Métodos para encontrar pruebas

8.2.1. Prueba de razón de verosimilitud

Si X_1, \dots, X_n es una muestra aleatoria de una población con función de probabilidad o función de densidad $f(x|\theta)$ entonces la función de verosimilitud es:

$$L(\theta|x_1, \dots, x_n) = L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$$

Definición 8.2.1 La prueba estadística de razón de verosimilitud para probar $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ es:

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}$$

donde Θ denota el espacio paramétrico completo. Una *prueba de razón de verosimilitud* es aquella con región de rechazo de la forma $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$, donde $0 \leq c \leq 1$.

Suponga que $\hat{\theta}$ es el estimador de máxima verosimilitud de θ sin restricción. Considerar además que $\hat{\theta}_0$ es el estimador de máxima verosimilitud de θ asumiendo que el espacio paramétrico restringido es Θ_0 , es decir $\hat{\theta}_0$ es el valor de $\theta \in \Theta_0$ que maximiza $L(\theta|\mathbf{x})$. Entonces, la prueba de razón de verosimilitud es:

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$$

Ejemplo 8.2.1 Sea X_1, \dots, X_n una muestra aleatoria de una población $\mathcal{N}(\theta, 1)$. Considerar la hipótesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ donde θ_0 es una constante especificada a priori por el experimentador. Como existe un solo valor de θ especificado por H_0 el numerador de $\lambda(\mathbf{x})$ es $L(\theta_0|\mathbf{x})$. El estimador de máxima verosimilitud no restringido de θ es \bar{x} , entonces la prueba de razón de verosimilitud es:

$$\begin{aligned} \lambda(\mathbf{x}) &= \frac{(2\pi)^{-n/2} \exp \{-\sum_{i=1}^n (x_i - \theta_0)^2 / 2\}}{(2\pi)^{-n/2} \exp \{-\sum_{i=1}^n (x_i - \bar{x})^2 / 2\}} \\ &= \exp \left\{ \frac{1}{2} \left(-\sum_{i=1}^n (x_i - \theta_0)^2 + \sum_{i=1}^n (x_i - \bar{x})^2 \right) \right\} \end{aligned}$$

además:

$$\sum_{i=1}^n (x_i - \theta_0)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \theta_0)^2$$

Luego,

$$\lambda(\mathbf{x}) = \exp \left\{ -n(\bar{x} - \theta_0)^2 / 2 \right\}$$

La región de rechazo $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$ se puede escribir como:

$$\left\{ \mathbf{x} : |\bar{x} - \theta_0| \geq \sqrt{-2 \log c / n} \right\}$$

8.3. Métodos para evaluar pruebas

8.3.1. Probabilidades de error y potencia de prueba

Una prueba de hipótesis para $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$ puede conducir a dos tipos de error.

	Decisión	
Hipótesis verdadera	No se rechaza H_0	Se rechaza H_0
H_0	Decisión correcta	Error tipo I
H_1	Error tipo II	Decisión correcta

Suponga que R denota la región de rechazo de una prueba, entonces:

$$\Pr_{\theta}(X \in R) = \begin{cases} \Pr(\text{Cometer error tipo I}) & \text{si } \theta \in \Theta_0 \\ 1 - \Pr(\text{Cometer error tipo II}) & \text{si } \theta \in \Theta_0^c \end{cases}$$

Definición 8.3.1 La *función potencia* de una prueba de hipótesis con región de rechazo R es la función de θ definida por $\beta(\theta) = \Pr_{\theta}(X \in R)$.

Ejemplo 8.3.1 Sea X_1, \dots, X_n una muestra aleatoria de una población $\mathcal{N}(\mu, \sigma^2)$, σ^2 conocido. Una prueba de razón de verosimilitud para $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ rechaza H_0 si $(\bar{x} - \theta_0)/(\sigma/\sqrt{n}) > c$. La constante c puede

ser cualquier número positivo. La función potencia para esta prueba es:

$$\begin{aligned}\beta(\theta) &= \Pr_{\theta} \left(\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} > c \right) \\ &= \Pr_{\theta} \left(\frac{\bar{X} - \theta}{\sigma/\sqrt{n}} > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right) \\ &= \Pr_{\theta} \left(Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)\end{aligned}$$

Conforme θ se incrementa desde $-\infty$ a ∞ es fácil ver que esta probabilidad se incrementa de cero a uno, es decir que $\beta(\theta)$ es una función creciente de θ con:

$$\lim_{\theta \rightarrow -\infty} \beta(\theta) = 0, \quad \lim_{\theta \rightarrow \infty} \beta(\theta) = 1 \quad \text{y} \quad \beta(\theta_0) = \alpha \quad \text{si} \quad \Pr(Z > c) = \alpha$$

donde α es la probabilidad de cometer error tipo I o nivel de significación de la prueba. El gráfico de $\beta(\theta)$ para $c = 1,28$ se muestra a continuación.

8.4. Pruebas óptimas para hipótesis compuestas

8.4.1. Familias exponenciales

Teorema 8.4.1 Sea X_1, \dots, X_n una muestra aleatoria de una función de probabilidad o densidad de la forma $f(x|\theta) = h(x)c(\theta) \exp\{w(\theta)t(x)\}$. Si $T(\mathbf{x}) = \sum_{i=1}^n t(x_i)$ y se desea probar $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

a. Si $w(\theta)$ es monótona creciente entonces:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } T(\mathbf{x}) \geq c \\ 0 & \text{otro modo} \end{cases}$$

es una prueba uniformemente más poderosa.

b. Si $w(\theta)$ es monótona decreciente entonces:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } T(\mathbf{x}) \leq c \\ 0 & \text{otro modo} \end{cases}$$

es una prueba uniformemente más poderosa.

Teorema 8.4.2 Sea X_1, \dots, X_n una muestra aleatoria de una función de probabilidad o densidad de la forma $f(x|\theta) = h(x)c(\theta) \exp\{w(\theta)t(x)\}$. Si $T(\mathbf{x}) = \sum_{i=1}^n t(x_i)$ y se desea probar $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$.

a. Si $w(\theta)$ es monótona creciente entonces:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } T(\mathbf{x}) \leq c \\ 0 & \text{otro modo} \end{cases}$$

es una prueba uniformemente más poderosa.

b. Si $w(\theta)$ es monótona decreciente entonces:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } T(\mathbf{x}) \geq c \\ 0 & \text{otro modo} \end{cases}$$

es una prueba uniformemente más poderosa.

Ejemplo 8.4.1 Sea X_1, \dots, X_5 una muestra aleatoria a partir de la distribución $\mathcal{N}(2, \theta)$. Hallar la prueba uniformemente más poderosa de tamaño $\alpha = 0,05$ para probar las hipótesis $H_0 : \theta \leq 0,8$ versus $H_1 : \theta > 0,8$.

8.4.2. Familias con la propiedad de razón de verosimilitud monótona

Definición 8.4.1 Una familia de densidades tiene la propiedad de *razón de verosimilitud monótona* si para todo $\theta_1 < \theta_2$:

$$\psi = \frac{\prod_{i=1}^n f(x_i|\theta_1)}{\prod_{i=1}^n f(x_i|\theta_2)}$$

es una función creciente o decreciente en $T(\mathbf{x})$.

Teorema 8.4.3 Sea X_1, \dots, X_n una muestra aleatoria de una función de densidad que pertenece a una familia con la propiedad de razón de verosimilitud monótona en $T(\mathbf{x})$ y se desea probar $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

a. Si la razón de verosimilitud monótona es creciente en $T(\mathbf{x})$ entonces:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } T(\mathbf{x}) \leq c \\ 0 & \text{otro modo} \end{cases}$$

es una prueba uniformemente más poderosa.

- b. Si la razón de verosimilitud monótona es decreciente en $T(\mathbf{x})$ entonces:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } T(\mathbf{x}) \geq c \\ 0 & \text{otro modo} \end{cases}$$

es una prueba uniformemente más poderosa.

Teorema 8.4.4 Sea X_1, \dots, X_n una muestra aleatoria de una función de densidad que pertenece a una familia con la propiedad de razón de verosimilitud monótona en $T(\mathbf{x})$ y se desea probar $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$.

- a. Si la razón de verosimilitud monótona es creciente en $T(\mathbf{x})$ entonces:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } T(\mathbf{x}) \geq c \\ 0 & \text{otro modo} \end{cases}$$

es una prueba uniformemente más poderosa.

- b. Si la razón de verosimilitud monótona es decreciente en $T(\mathbf{x})$ entonces:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } T(\mathbf{x}) \leq c \\ 0 & \text{otro modo} \end{cases}$$

es una prueba uniformemente más poderosa.

Ejemplo 8.4.2 Sea X_1, \dots, X_n una muestra aleatoria a partir de la distribución $f(x|\theta) = 2x/\theta^2$, $0 < x < \theta$. Hallar la prueba uniformemente más poderosa de tamaño α para probar las hipótesis $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$.

8.4.3. Distribución asintótica de la prueba de razón de verosimilitud

Teorema 8.4.5 Sea X_1, \dots, X_n una muestra aleatoria de una función de probabilidad o densidad $f(x|\theta)$. Bajo ciertas condiciones de regularidad sobre el modelo $f(x|\theta)$ y si $\theta \in \Theta_0$ entonces la distribución del estadístico $-2 \log \lambda(\mathbf{X})$ converge a la distribución χ^2 conforme $n \rightarrow \infty$. Los grados de libertad de la distribución límite se obtienen como la diferencia entre el número de parámetros libres especificados por $\theta \in \Theta_0$ y el número de parámetros libres especificados por $\theta \in \Theta$.

Ejemplo 8.4.3 Sea X_1, \dots, X_n una muestra aleatoria a partir de la distribución $\mathcal{N}(\theta, 1)$. Use la prueba de razón de verosimilitud monótona de tamaño $\alpha = 0,05$ para probar las hipótesis $H_0 : \theta = 3$ versus $H_1 : \theta \neq 3$.

Capítulo 9

Estimación por intervalos

9.1. Introducción

Definición 9.1.1 Una *estimación por intervalo* de un parámetro θ es algún par de funciones de la muestra, $L(x_1, \dots, x_n)$ y $U(x_1, \dots, x_n)$, que satisfacen $L(\mathbf{x}) \leq U(\mathbf{x})$ para todo $\mathbf{x} \in \mathcal{X}$. El intervalo aleatorio $[L(\mathbf{X}), U(\mathbf{X})]$ es llamado un *estimador por intervalo*.

Ejemplo 9.1.1 Para una muestra X_1, \dots, X_n de la distribución $N(\mu, 1)$ un estimador por intervalo de μ es $[\bar{X} - 1, \bar{X} + 1]$.

Definición 9.1.2 Sea $[L(\mathbf{X}), U(\mathbf{X})]$ un estimador por intervalo de θ , la *probabilidad de cobertura* de $[L(\mathbf{X}), U(\mathbf{X})]$ es la probabilidad que el intervalo aleatorio cubra al parámetro θ , es decir, $\Pr_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$.

Ejemplo 9.1.2 En el ejemplo anterior la probabilidad que μ sea cubierto por $[\bar{X} - 1, \bar{X} + 1]$ es 0,9544.

Definición 9.1.3 Sea $[L(\mathbf{X}), U(\mathbf{X})]$ un estimador por intervalo para θ , el coeficiente de confianza es el ínfimo de las probabilidades de cobertura, es decir $\inf_\theta \Pr_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$.

Los estimadores de intervalo junto con una medida de confianza, usualmente un coeficiente de confianza, son conocidos como *intervalos de confianza*.

Ejemplo 9.1.3 Sea X_1, \dots, X_n una muestra aleatoria de una población unifome($0, \theta$) y sea $X_{(n)}$ el máximo. Se está interesado en un estadístico por intervalo de θ . Si consideramos los siguientes candidatos:

$$[aX_{(n)}, bX_{(n)}] \quad \text{y} \quad [X_{(n)} + c, X_{(n)} + d]$$

$1 \leq a < b$ $0 \leq c < d$

donde a , b , c y d son constantes. Notar que θ es necesariamente mayor que $X_{(n)}$. Para el primer intervalo se tiene:

$$\begin{aligned} \Pr_{\theta}(\theta \in [aX_{(n)}, bX_{(n)}]) &= \Pr_{\theta}(aX_{(n)} \leq \theta \leq bX_{(n)}) \\ &= \Pr\left(\frac{1}{b} \leq \frac{X_{(n)}}{\theta} \leq \frac{1}{a}\right) \\ &= \Pr\left(\frac{1}{b} \leq T \leq \frac{1}{a}\right) \\ &= \left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n \end{aligned}$$

y no depende del parámetro. Para el segundo intervalo:

$$\begin{aligned} \Pr_{\theta}(\theta \in [X_{(n)} + c, X_{(n)} + d]) &= \Pr_{\theta}(X_{(n)} + c \leq \theta \leq X_{(n)} + d) \\ &= \Pr\left(1 - \frac{d}{\theta} \leq T \leq 1 - \frac{c}{\theta}\right) \\ &= \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n \end{aligned}$$

y depende del parámetro.

9.2. Métodos para encontrar estimadores de intervalos

9.2.1. Invirtiendo una prueba estadística

Ejemplo 9.2.1 Sean X_1, \dots, X_n independientes e idénticamente distribuidas según $N(\mu, \sigma^2)$. Considere las hipótesis $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. Para un nivel α , se tiene la región de rechazo $\{\mathbf{x} : |\bar{x} - \mu_0| > z_{1-\alpha/2}\sigma/\sqrt{n}\}$. Notar que H_0 no se rechaza si $|\bar{x} - \mu_0| \leq z_{1-\alpha/2}\sigma/\sqrt{n}$, o equivalentemente:

$$\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Como la prueba tiene tamaño α , esto significa que $\Pr(H_0 \text{ se rechaza} / \mu = \mu_0) = \alpha$, o visto de otra forma $\Pr(\text{No rechazar } H_0 / \mu = \mu_0) = 1 - \alpha$. Luego:

$$\Pr \left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \middle| \mu = \mu_0 \right) = 1 - \alpha$$

Pero lo anterior es verdadero para todo μ_0 , entonces:

$$\Pr_{\mu} \left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

Teorema 9.2.1 Para todo $\theta_0 \in \Theta$, sea $A(\theta_0)$ la región de no rechazo para una prueba con nivel α de $H_0 : \theta = \theta_0$. Para cada $\mathbf{x} \in \mathcal{X}$ se define el conjunto $C(\mathbf{x})$ en el espacio paramétrico por:

$$C(\mathbf{x}) = \{\theta_0 : \mathbf{x} \in A(\theta_0)\} \quad (9.2.1)$$

Entonces el conjunto aleatorio $C(\mathbf{X})$ es un conjunto de confianza $1 - \alpha$. Inversamente, sea $C(\mathbf{X})$ un conjunto de confianza $1 - \alpha$. Para todo $\theta_0 \in \Theta$, se define:

$$A(\theta_0) = \{\mathbf{x} : \theta_0 \in C(\mathbf{x})\}$$

Entonces $A(\theta_0)$ es la región de no rechazo de una prueba a un nivel α de $H_0 : \theta = \theta_0$.

Ejemplo 9.2.2 Suponga que se desea un intervalo de confianza para la media λ de una población exponencial usando la prueba $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda \neq \lambda_0$ a un nivel α :

$$\lambda(\mathbf{x}) = \frac{\frac{1}{\lambda_0^n} e^{-\sum x_i / \lambda_0}}{\sup_{\lambda} \frac{1}{\lambda^n} e^{-\sum x_i / \lambda}} = \frac{\frac{1}{\lambda_0^n} e^{-\sum x_i / \lambda_0}}{\left(\frac{1}{(\sum x_i / n)^n} e^{-n} \right)} = \left(\frac{\sum x_i}{n \lambda_0} \right)^n e^n e^{-\sum x_i / \lambda_0}$$

Para λ_0 fijo la región de no rechazo es:

$$A(\lambda_0) = \left\{ \mathbf{x} : \left(\frac{\sum x_i}{n \lambda_0} \right)^n e^{-\sum x_i / \lambda_0} \geq c \right\}$$

donde c es la constante que satisface $\Pr_{\lambda_0}(\mathbf{x} \in A(\lambda_0)) = 1 - \alpha$. Invirtiendo la región de no rechazo se obtiene el conjunto de confianza $1 - \alpha$:

$$C(\mathbf{x}) = \left\{ \lambda : \left(\frac{\sum x_i}{\lambda} \right)^n e^{-\sum x_i/\lambda} \geq c \right\}$$

9.2.2. Cantidades pivotaes

Definición 9.2.1 Una variable aleatoria $Q(\mathbf{X}, \theta) = Q(X_1, \dots, X_n, \theta)$ es una *cantidad pivotal*, o pivote, si la distribución de $Q(\mathbf{X}, \theta)$ es independiente de todo parámetro. Es decir, si $\mathbf{X} \sim F(\mathbf{x}/\theta)$ entonces $Q(\mathbf{X}, \theta)$ tiene la misma distribución para todos los valores de θ .

Ejemplo 9.2.3 En los casos de las familias de locación y escala existen muchas cantidades pivotaes

Forma	Tipo	Cantidad pivotal
$f(x - \mu)$	Locación	$\bar{X} - \mu$
$\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$	Escala	$\frac{\bar{X}}{\sigma}$
$\frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$	Locación-escala	$\frac{\bar{X} - \mu}{S}$

Ejemplo 9.2.4 Suponga que X_1, \dots, X_n son independientes e idénticamente distribuidos según la exponencial(λ). Entonces $T = \sum X_i$ es una estadística suficiente para λ y $T \sim \text{gamma}(n, \lambda)$, luego $Q(T, \lambda) = 2T/\lambda \sim \chi_{2n}^2$ es un pivote. Recordar además que la distribución gamma es una familia de escala.

Ejemplo 9.2.5 En el ejemplo 9.2.2 se obtuvo un intervalo de confianza para la media invirtiendo la prueba de nivel α , $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda \neq \lambda_0$. Si se tiene una muestra aleatoria X_1, \dots, X_n y se define $T = \sum X_i$ y $Q(T, \lambda) = 2T/\lambda \sim \chi_{2n}^2$, pueden escogerse las constantes a y b que satisfacen $\Pr(a \leq \chi_{2n}^2 \leq b) = 1 - \alpha$, entonces:

$$\begin{aligned} \Pr(a \leq Q(T, \lambda) \leq b) &= \Pr\left(a \leq \frac{2t}{\lambda} \leq b\right) \\ &= \Pr\left(\frac{2t}{b} \leq \lambda \leq \frac{2t}{a}\right) \\ &= 1 - \alpha \end{aligned}$$

Ejemplo 9.2.6 Si X_1, \dots, X_n son independientes e idénticamente distribuidas según $N(\mu, \sigma^2)$, entonces $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ es un pivote cuando σ^2 es conocido y puede utilizarse para calcular un intervalo de confianza para μ :

$$\Pr\left(-a \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq a\right) = \Pr(-a \leq Z \leq a)$$

entonces:

$$\left\{ \mu : \bar{x} - a \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + a \frac{\sigma}{\sqrt{n}} \right\}$$

9.2.3. Garantizado un intervalo

Teorema 9.2.2 Sea T una estadística con función de distribución acumulada $F_T(t/\theta)$ y $0 < \alpha < 1$ una constante fija. Suponga que para cada $t \in T$, las funciones $\theta_L(t)$ y $\theta_U(t)$ se definen como:

- a. Si $F_T(t/\theta)$ es una función decreciente de θ para cada t , se define $\theta_L(t)$ y $\theta_U(t)$ por:

$$F_T(t/\theta_U(t)) = \alpha/2, \quad F_T(t/\theta_L(t)) = 1 - \alpha/2$$

- b. Si $F_T(t/\theta)$ es una función creciente de θ para cada t , se define $\theta_L(t)$ y $\theta_U(t)$ por:

$$F_T(t/\theta_U(t)) = 1 - \alpha/2, \quad F_T(t/\theta_L(t)) = \alpha/2$$

Entonces, el intervalo aleatorio $[\theta_L(T), \theta_U(T)]$ es un intervalo de confianza $1 - \alpha$ para θ .

Ejemplo 9.2.7 Si X_1, \dots, X_n son independientes con función de densidad $f(x_i/\theta) = e^{-(x-\theta)} I_{[\theta, \infty)}(x)$, entonces $T = X_{(1)}$ es una estadística suficiente para θ con función de densidad:

$$f_T(t/\theta) = ne^{-n(t-\theta)} I_{[\theta, \infty)}(t)$$

cuya función de distribución acumulada:

$$F_T(t/\theta) = 1 - e^{-n(t-\theta)}$$

es una función decreciente de θ . Entonces, fijado α y definidos $\theta_L(t)$ y $\theta_U(t)$ que satisfacen las ecuaciones:

$$F_T(t/\theta_U(t)) = 1 - e^{-n(t-\theta_U(t))} = \frac{\alpha}{2}, \quad F_T(t/\theta_L(t)) = 1 - e^{-n(t-\theta_L(t))} = 1 - \alpha/2$$

cuyas soluciones son:

$$\theta_U(t) = t + \frac{1}{n} \log \left(1 - \frac{\alpha}{2} \right), \quad \theta_L(t) = t + \frac{1}{n} \log \left(\frac{\alpha}{2} \right)$$

Luego:

$$C(X_{(1)}) = \left\{ \theta : X_{(1)} + \frac{1}{n} \log \left(\frac{\alpha}{2} \right) \leq \theta \leq X_{(1)} + \frac{1}{n} \log \left(1 - \frac{\alpha}{2} \right) \right\}$$

es un intervalo de confianza $(1 - \alpha)$ para θ .

Teorema 9.2.3 Sea T una estadística discreta con función de probabilidad acumulada $F_T(t/\theta) = \Pr(T \leq t/\theta)$ y $0 < \alpha < 1$ un valor fijo. Suponga que para todo $t \in T$, $\theta_L(t)$ y $\theta_U(t)$ pueden definirse como sigue:

- a. Si $F_T(t/\theta)$ es una función decreciente de θ para cada t , se define $\theta_L(t)$ y $\theta_U(t)$ por:

$$\Pr(T \leq t/\theta_U(t)) = \alpha/2, \quad \Pr(T \geq t/\theta_L(t)) = \alpha/2$$

- b. Si $F_T(t/\theta)$ es una función creciente de θ para cada t , se define $\theta_L(t)$ y $\theta_U(t)$ por:

$$\Pr(T \geq t/\theta_U(t)) = \alpha/2, \quad \Pr(T \leq t/\theta_L(t)) = \alpha/2$$

Entonces, el intervalo aleatorio $[\theta_L(T), \theta_U(T)]$ es un intervalo de confianza $1 - \alpha$ para θ .

Ejemplo 9.2.8 Sea X_1, \dots, X_n una muestra aleatoria de una población Poisson con parámetro λ y se define la estadística suficiente $Y = \sum X_i$ cuya distribución es $\text{Poisson}(n\lambda)$. Aplicando el método anterior si se observa $Y = y_0$ esto conduce a resolver las ecuaciones:

$$\sum_{k=0}^{y_0} e^{-n\lambda} \frac{(n\lambda)^k}{k!} = \frac{\alpha}{2} \quad \text{y} \quad \sum_{k=y_0}^{\infty} e^{-n\lambda} \frac{(n\lambda)^k}{k!} = \frac{\alpha}{2}$$

Recordando la identidad 3.2.7 entre las familias Poisson y gamma se tiene:

$$\frac{\alpha}{2} = \sum_{k=0}^{y_0} e^{-n\lambda} \frac{(n\lambda)^k}{k!} = \Pr(Y \leq y_0/\lambda) = \Pr(\chi_{2(y_0+1)}^2 > 2n\lambda)$$

cuya solución es $\lambda = \frac{1}{2n} \chi_{2(y_0+1), \alpha/2}^2$. De manera similar para la segunda ecuación se obtiene:

$$\frac{\alpha}{2} = \sum_{k=y_0}^{\infty} e^{-n\lambda} \frac{(n\lambda)^k}{k!} = \Pr(Y \leq y_0/\lambda) = \Pr(\chi_{2(y_0+1)}^2 < 2n\lambda)$$

Finalmente, el intervalo de confianza $1 - \alpha$ para λ es:

$$\left\{ \lambda : \frac{1}{2n} \chi_{2y_0, 1-\alpha/2}^2 \leq \lambda \leq \frac{1}{2n} \chi_{2(y_0+1), \alpha/2}^2 \right\}$$

Si $y_0 = 0$, se define $\chi_{0, 1-\alpha/2}^2 = 0$.

9.3. Métodos de evaluación de estimadores por intervalos

9.3.1. Tamaño y probabilidad de cobertura

Ejemplo 9.3.1 Sean X_1, \dots, X_n independientes y distribuidas según $N(\mu, \sigma^2)$ donde σ^2 es conocida. Usando el método de la sección 9.2.2 y además:

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

es un pivote con distribución normal estándar. Considerando a y b que satisfacen:

$$\Pr(a \leq Z \leq b) = 1 - \alpha$$

permite obtener el intervalo de confianza $1 - \alpha$

$$\left\{ \mu : \bar{x} - b \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} - a \frac{\sigma}{\sqrt{n}} \right\}$$

¿Qué elección a y b es la mejor? ¿que elección de a y b minimiza la longitud del intervalo de confianza manteniendo la cobertura $1 - \alpha$?

Teorema 9.3.1 Sea $f(x)$ una función de densidad unimodal. Si el intervalo $[a, b]$ satisface:

- a. $\int_a^b f(x)dx = 1 - \alpha$.
- b. $f(a) = f(b) > 0$, y
- c. $a \leq x^* \leq b$, donde x^* es una moda de $f(x)$.

entonces $[a, b]$ es el más pequeño entre todos los intervalos con cobertura $1 - \alpha$.

Ejemplo 9.3.2 Para intervalos de la distribución normal basados en la cantidad pivotal $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ se sabe que el intervalo de confianza $1 - \alpha$ de longitud más pequeña es de la forma:

$$\bar{x} - b \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} - a \frac{s}{\sqrt{n}}$$

La longitud del intervalo es una función de s :

$$\text{Longitud}(s) = (b - a) \frac{s}{\sqrt{n}}$$

Aplicando el teorema 9.3.1 se llega a que $a = -t_{n-1; \alpha/2}$ y $b = t_{n-1; \alpha/2}$ permite obtener el intervalo óptimo.

9.4. Otras consideraciones

9.4.1. Intervalos aproximados por máxima verosimilitud

Si X_1, \dots, X_n son independientes distribuidas según $f(x/\theta)$ y $\hat{\theta}$ es el estimador de máxima verosimilitud para θ , entonces de 7.4.1 la varianza de una función $h(\hat{\theta})$ puede ser aproximada por:

$$\widehat{\text{Var}}(h(\hat{\theta})/\theta) \approx \frac{[h'(\theta)]^2 |_{\theta=\hat{\theta}}}{-\frac{\partial^2}{\partial \theta^2} \log L(\theta/\mathbf{x}) |_{\theta=\hat{\theta}}}$$

Luego, para un valor de θ arbitrario pero fijo y bajo condiciones generales de regularidad se tiene:

$$\frac{h(\hat{\theta}) - h(\theta)}{\sqrt{\widehat{\text{Var}}(h(\hat{\theta})/\theta)}} \rightarrow N(0, 1)$$

lo cual permite obtener el intervalo aproximado de confianza:

$$h(\hat{\theta}) - z_{\alpha/2} \sqrt{\widehat{\text{Var}}(h(\hat{\theta})/\theta)} \leq h(\theta) < h(\hat{\theta}) + z_{\alpha/2} \sqrt{\widehat{\text{Var}}(h(\hat{\theta})/\theta)}$$

Ejemplo 9.4.1 Se tiene una muestra aleatoria X_1, \dots, X_n de una población Bernoulli(p). Si se desea estimar la razón de odds $p/(1-p)$ puede utilizarse $\hat{p}/(1-\hat{p})$ donde \hat{p} es el estimador de máxima verosimilitud. Luego:

$$\widehat{\text{Var}}\left(\frac{\hat{p}}{1-\hat{p}}\right) \approx \frac{[h'(\theta)]^2 |_{\theta=\hat{\theta}}}{-\frac{\partial^2}{\partial \theta^2} \log L(\theta/\mathbf{x}) |_{\theta=\hat{\theta}}} = \frac{\left[\frac{1}{(1-\hat{p})^2}\right]^2}{\frac{n}{\hat{p}(1-\hat{p})}} = \frac{\hat{p}}{n(1-\hat{p})^3}$$

Finalmente, se puede construir el intervalo de confianza aproximado:

$$\frac{\hat{p}}{1-\hat{p}} - z_{\alpha/2} \sqrt{\widehat{\text{Var}}\left(\frac{\hat{p}}{1-\hat{p}}\right)} \leq \frac{p}{1-p} \leq \frac{\hat{p}}{1-\hat{p}} + z_{\alpha/2} \sqrt{\widehat{\text{Var}}\left(\frac{\hat{p}}{1-\hat{p}}\right)}$$

9.4.2. Otros intervalos aproximados

Si se tienen las estadísticas W, V y un parámetro θ tal que cuando $n \rightarrow \infty$,

$$\frac{W - \theta}{V} \rightarrow N(0, 1)$$

entonces se puede construir un intervalo de confianza aproximado para θ por:

$$W - z_{\alpha/2} V \leq \theta \leq W + z_{\alpha/2} V$$