

Capítulo 6 Funciones de variables aleatorias

6.1 Introducción

El objetivo de la estadística es hacer inferencias acerca de una población con base en información contenida en una muestra tomada de esa población. Cualquier inferencia verdaderamente útil debe ser acompañada por una medida de bondad asociada. Frecuentemente se utiliza el *error de estimación*, que es la diferencia entre la estimación y el parámetro estimado. Como se necesitan ocupar n variables aleatorias para calcular un estimador entonces este estimador también es una variable aleatoria. Con esto se deduce que se puede determinar la distribución de probabilidad para una función de n variables aleatorias Y_1, Y_2, \dots, Y_n . Esto se puede hacer calculando la distribución de probabilidad conjunta para estas variables. Para este capítulo supondremos que las variables fueron tomadas de una muestra con una población lo suficientemente grande para que estas sean independientes. Entonces para el caso discreto la función de probabilidad conjunta para las variables Y_1, Y_2, \dots, Y_n , está dada por

$$p(y_1, y_2, \dots, y_n) = p(y_1)p(y_2) \cdots p(y_n)$$

En el caso continuo, la función de densidad conjunta es

$$f(y_1, y_2, \dots, y_n) = f(y_1)f(y_2) \cdots f(y_n)$$

En las siguientes tres secciones se verán métodos para determinar la distribución de probabilidad para una función de variables aleatorias y un cuarto método para determinar la distribución *conjunta* de varias funciones de variables aleatorias.

6.3 Método de las funciones de distribución

Este método se emplea generalmente cuando las Y tienen distribuciones continuas. Si Y tiene una función de densidad de probabilidad $f_Y(y)$ y si U es alguna función de Y , entonces podemos calcular $F_U(u) = P(U \leq u)$ directamente al integrar $f_Y(y)$ en la región para la cual $U \leq u$. La función de densidad de probabilidad para U se encuentra al derivar $F_U(u)$. Entonces se puede resumir el método de la siguiente manera.

Resumen del método de las funciones de distribución

Sea U una función de las variables aleatorias Y_1, Y_2, \dots, Y_n .

1. Localice la región $U = u$ del espacio (y_1, y_2, \dots, y_n) .
2. Localice la región $U \leq u$.
3. Determine $F_U(u) = P(U \leq u)$ al integrar $f_Y(y_1, y_2, \dots, y_n)$ sobre la región $U \leq u$.
4. Determine la función de densidad $f_U(u)$ al derivar $F_U(u)$. Por tanto, $f_U(u) = \frac{dF_U(u)}{du}$.

Para ilustrar, considere el caso $U = h(Y) = Y^2$, donde Y es una variable aleatoria continua con función de distribución $F_Y(y)$ y función de densidad $f_Y(y)$. Si $u \leq 0$, $F_U(u) = P(U \leq u) = P(Y^2 \leq u) = 0$ y para $u > 0$,

$$\begin{aligned} F_U(u) &= P(U \leq u) = P(Y^2 \leq u) \\ &= P(-\sqrt{u} \leq Y \leq \sqrt{u}) \\ &= \int_{-\sqrt{u}}^{\sqrt{u}} f_Y(y) dy = F_Y(\sqrt{u}) - F_Y(-\sqrt{u}) \end{aligned}$$

En general,

$$F_U(u) = \begin{cases} F_Y(\sqrt{u}) - F_Y(-\sqrt{u}), & u > 0 \\ 0, & \text{en cualquier otro punto} \end{cases}$$

Al derivar con respecto a u , vemos que

$$f_U(u) = \begin{cases} f_Y(\sqrt{u}) \left(\frac{1}{2\sqrt{u}} \right) + f_Y(-\sqrt{u}) \left(\frac{1}{2\sqrt{u}} \right), & u > 0 \\ 0, & \text{en cualquier otro punto} \end{cases}$$

o bien, simplemente,

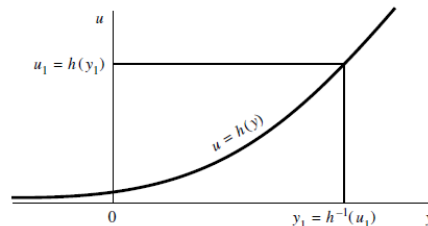
$$f_U(u) = \begin{cases} \left(\frac{1}{2\sqrt{u}} \right) [f_Y(\sqrt{u}) + f_Y(-\sqrt{u})], & u > 0 \\ 0, & \text{en cualquier otro punto} \end{cases}$$

En algunos casos es posible encontrar una transformación que cuando se aplica a una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$, resulta en una variable aleatoria con alguna otra función de distribución específica, por ejemplo $F(y)$.

6.4 Método de las transformaciones

El método de las transformaciones, es una consecuencia del método de función de distribución de la sección anterior. Mediante el método de las funciones de distribución podemos llegar a un método simple para formular la función de densidad de $U = h(Y)$ sea decreciente o creciente.

Suponga que $h(y)$ es una función creciente de y y que $U = h(Y)$, donde Y tiene función de densidad $f_Y(y)$. Entonces $h^{-1}(u)$ es una función creciente de u : si $u_1 < u_2$, entonces $h^{-1}(u_1) = y_1 < y_2 = h^{-1}(u_2)$. En la siguiente figura vemos que el conjunto de puntos y tales que $h(y) \leq u_1$ es precisamente igual que el conjunto de puntos y tales que $y \leq h^{-1}(u_1)$.



Por lo tanto,

$$\begin{aligned} P(U \leq u) &= P[h(Y) \leq u] = P\{h^{-1}[h(Y)] \leq h^{-1}(u)\} \\ &= P[Y \leq h^{-1}(u)] \end{aligned}$$

o bien,

$$F_U(u) = F_Y[h^{-1}(u)]$$

Entonces derivando con respecto a u , tenemos

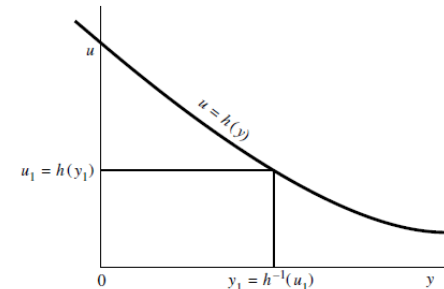
$$f_U(u) = \frac{dF_U(u)}{du} = \frac{dF_Y[h^{-1}(u)]}{du} = f_Y[h^{-1}(u)] \frac{d[h^{-1}(u)]}{du}$$

Para simplificar la notación, escribiremos dh^{-1}/du en lugar de $d[h^{-1}(u)]/du$ y

$$f_U(u) = f_Y[h^{-1}(u)] \frac{dh^{-1}}{du}$$

Para determinar $f_U(u)$, despeje y en término de u ; esto es, encuentre $y = h^{-1}(u)$ y sustituya esta expresión en $f_Y(y)$. Finalmente multiplique esta cantidad por dh^{-1}/du .

Si $h(y)$ es una función decreciente de y , entonces $h^{-1}(u)$ es una función decreciente de u . Esto es, si $u_1 < u_2$, entonces $h^{-1}(u_1) = y_1 > y_2 = h^{-1}(u_2)$. Al igual que el caso creciente, de la siguiente figura se puede observar como el conjunto de puntos y tales que $h(y) \leq u_1$ es el mismo conjunto de puntos tales que $y \geq h^{-1}(u_1)$.



Se deduce que, para $U = h(Y)$,

$$P(U \leq u) = P[Y \geq h^{-1}(u)] \quad \text{o} \quad F_U(u) = 1 - F_Y[h^{-1}(u)]$$

Si derivamos con respecto a u , obtenemos

$$f_U(u) = -f_Y[h^{-1}(u)] \frac{d[h^{-1}(u)]}{du}$$

Se de nuevo empleamos la notación simplificada de dh^{-1}/du en lugar de $d[h^{-1}(u)]/du$ y recordamos que dh^{-1}/du es negativa porque $h^{-1}(u)$ es una función decreciente de u , la densidad de U es

$$f_U(u) = f_Y[h^{-1}(u)] \left| \frac{dh^{-1}}{du} \right|$$

En realidad, no es necesario que $h(y)$ sea creciente o decreciente para todos los valores de y . La función $h(\cdot)$ sólo necesita ser creciente o decreciente para los valores de y tales que $f_Y(y) > 0$. El conjunto de puntos $\{y : f_Y(y) > 0\}$ se denomina *soporte* de la densidad $f_Y(y)$. Si $y = h^{-1}(u)$ no es el soporte de la densidad, entonces $f_Y[h^{-1}(u)] = 0$. En otras palabras se tiene,

Sea Y la función de densidad de probabilidad $f_Y(y)$. Si $h(y)$ es creciente o decreciente para toda y tal que $f_Y(y) > 0$, entonces $U = h(Y)$ tiene función de densidad

$$f_U(u) = f_Y[h^{-1}(u)] \left| \frac{dh^{-1}}{du} \right|$$

$$\text{donde } \frac{dh^{-1}}{du} = \frac{d[h^{-1}(u)]}{du}$$

Se tiene que comprobar que $y = h^{-1}(u)$ es el soporte de la densidad antes de usar este método. De otro modo no se puede aplicar las técnicas vistas en este método. En resumen se tiene lo siguiente,

Resumen del método de las transformaciones

Sea $U = h(Y)$, donde $h(y)$ es una función creciente o decreciente de y para toda y tal que $f_Y(y) > 0$.

1. Deduzca la función inversa, $y = h^{-1}(u)$.
2. Evalúe $\frac{dh^{-1}}{du} = \frac{d[h^{-1}(u)]}{du}$
3. Encuentre $f_U(u)$ mediante

$$f_U = f_Y[h^{-1}(u)] \left| \frac{dh^{-1}}{du} \right|$$

6.5 Método de las funciones generadoras de momento

El método de las funciones generadoras de momento para determinar la distribución de probabilidad de una función de variables aleatorias Y_1, Y_2, \dots, Y_n está basada en el siguiente teorema de unicidad.

Teorema 6.1

Denotemos con $m_X(t)$ y $m_Y(t)$ las funciones generadoras de momento de variables aleatorias X y Y respectivamente. Si existen funciones generadoras de momento y $m_X(t) = m_Y(t)$ para todos los valores de t , entonces X y Y tienen la misma distribución de probabilidad.

Si U es una función de n variables aleatorias Y_1, Y_2, \dots, Y_n , el primer paso al usar el Teorema 6.1 es hallar la función generadora de momento de U . Una vez que se tiene esto se compara con las funciones generadoras de momento para variables aleatorias con distribuciones bien conocidas. Si $m_U(t)$ es idéntica a una de estas entonces poseen distribuciones de probabilidad idénticas. Se puede ilustrar este método con la siguiente demostración.

Demostración

Sea Y una variable aleatoria normalmente distribuida con media μ y varianza σ^2 . Entonces la variable aleatoria Z definida como

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

tiene una distribución *normal estándar*. Se desea demostrar que Z tiene una distribución normal con media 0 y varianza 1. Se tiene la información que $Y - \mu$ tiene función generadora de momento $e^{t^2 \sigma^2 / 2}$. Entonces usando el método explicado se tiene.

$$\begin{aligned} m_Z(t) &= E(e^{tZ}) = E\left[e^{(t/\sigma)(Y-\mu)}\right] = m_{(Y-\mu)}\left(\frac{t}{\sigma}\right) \\ &= e^{(t/\sigma)^2 (\sigma^2/2)} = e^{t^2/2}. \end{aligned}$$

Al comparar $m_Z(t)$ con la función generadora de momento de una variable aleatoria normal, veremos que Z debe estar normalmente distribuida con $E(Z) = 0$ y $V(Z) = 1$.

Este método es con frecuencia muy útil para hallar las distribuciones de sumas de variables aleatorias independientes.

Teorema 6.2

Sean Y_1, Y_2, \dots, Y_n variables aleatorias independientes con funciones generadoras de momento $m_{Y_1}, m_{Y_2}, \dots, m_{Y_n}$, respectivamente. Si $U = Y_1 + Y_2 + \dots + Y_n$, entonces

$$m_U(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times \dots \times m_{Y_n}(t)$$

Demostración

Sabemos que, como las variables aleatorias Y_1, Y_2, \dots, Y_n son independientes (vea el Teorema 5.9),

$$\begin{aligned} m_U(t) &= E\left[e^{t(Y_1 + \dots + Y_n)}\right] = E\left(e^{tY_1} e^{tY_2} \dots e^{tY_n}\right) \\ &= E(e^{tY_1}) \times E(e^{tY_2}) \times \dots \times E(e^{tY_n}) \end{aligned}$$

Entonces, por la definición de funciones generadoras de momentos,

$$m_U(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times \dots \times m_{Y_n}(t)$$

Teorema 6.3

Sean Y_1, Y_2, \dots, Y_n variables aleatorias independientes normalmente distribuidas con $E(Y_i) = \mu_i$ y $V(Y_i) = \sigma_i^2$, para $i = 1, 2, \dots, n$, y sean a_1, a_2, \dots, a_n constantes. Si

$$U = \sum_{i=1}^n a_i Y_i = a_1 Y_1 + a_2 Y_2 + \dots + a_n Y_n$$

entonces U es una variable aleatoria normalmente distribuida con

$$E(U) = \sum_{i=1}^n a_i \mu_i = a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n$$

y

$$V(U) = \sum_{i=1}^n a_i^2 \sigma_i^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2$$

Demostración

Cuando Y_i está normalmente distribuida con media μ_i y varianza σ_i^2 , Y_i tiene función generadora de momento dada por

$$m_{Y_i}(t) = \exp\left(\mu_i t + \frac{\sigma_i^2 t^2}{2}\right)$$

[Recuerde que $\exp(\cdot)$ es una forma más cómoda de escribir $e^{(\cdot)}$ cuando el término en el exponente es largo o complejo.] Por lo tanto, $a_i Y_i$ tiene función generadora de momento dada por

$$\begin{aligned} m_{a_i Y_i}(t) &= E(e^{t a_i Y_i}) = m_{Y_i}(a_i t) \\ &= \exp\left(\mu_i a_i t + \frac{a_i^2 \sigma_i^2 t^2}{2}\right) \end{aligned}$$

Debido a que las variables aleatorias Y_i son independientes, las variables aleatorias $a_i Y_i$ son independientes, para $i = 1, 2, \dots, n$ y el Teorema 6.2 implica que

$$\begin{aligned} m_U(t) &= m_{a_1 Y_1}(t) \times m_{a_2 Y_2}(t) \times \dots \times m_{a_n Y_n}(t) \\ &= \exp\left(\mu_1 a_1 t + \frac{a_1^2 \sigma_1^2 t^2}{2}\right) \times \dots \times \exp\left(\mu_n a_n t + \frac{a_n^2 \sigma_n^2 t^2}{2}\right) \\ &= \exp\left(t \sum_{i=1}^n a_i \mu_i + \frac{t^2}{2} \sum_{i=1}^n a_i^2 \sigma_i^2\right) \end{aligned}$$

Entonces, U tiene una distribución normal con media $\sum_{i=1}^n a_i \mu_i$ y varianza $\sum_{i=1}^n a_i^2 \sigma_i^2$.

Teorema 6.4

Sean Y_1, Y_2, \dots, Y_n definidas como en el Teorema 6.3 y definimos Z_i por

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i} \quad i = 1, 2, \dots, n$$

Entonces $\sum_{i=1}^n Z_i^2$ tiene una distribución χ^2 con n grados de libertad.

Demostración

Como Y_i está normalmente distribuida con media μ_i y varianza σ_i^2 , entonces Z_i tiene distribución normal con media 0 y varianza 1 como se vio en un resultado anterior. Del Ejemplo 6.11, entonces tenemos que Z_i^2 es una variable aleatoria con distribución χ^2 con 1 grado de libertad. En consecuencia,

$$m_{Z_i^2}(t) = (1 - 2t)^{-1/2}$$

y del Teorema 6.2, con $V = \sum_{i=1}^n Z_i^2$,

$$\begin{aligned} m_V(t) &= m_{Z_1^2}(t) \times m_{Z_2^2}(t) \times \dots \times m_{Z_n^2}(t) \\ &= (1 - 2t)^{-1/2} \times (1 - 2t)^{-1/2} \times \dots \times (1 - 2t)^{-1/2} \\ &= (1 - 2t)^{-n/2} \end{aligned}$$

Como las funciones generadoras de momento son únicas, V tiene una distribución χ^2 con n grados de libertad.

Resumen del método de las funciones generadoras de momento

Sea U una función de las variables aleatorias Y_1, Y_2, \dots, Y_n .

1. Encuentre la función generadora de momento para U , $m_U(t)$.
2. Compare $m_U(t)$ con otras funciones generadoras de momento bien conocidas. Si $m_U(t) = m_V(t)$ para todos los valores de t , el Teorema 6.1 implica que U y V tienen distribuciones idénticas.

Método de las transformaciones bivariadas

Suponga que Y_1 y Y_2 son variables aleatorias continuas con función de densidad conjunta $f_{Y_1, Y_2}(y_1, y_2)$ y que para toda (y_1, y_2) tal que $f_{Y_1, Y_2}(y_1, y_2) > 0$.

$$u_1 = h_1(y_1, y_2) \quad u_2 = h_2(y_1, y_2)$$

es una transformación uno a uno de (y_1, y_2) y (u_1, u_2) con inversa

$$y_1 = h_1^{-1}(u_1, u_2) \quad y_2 = h_2^{-1}(u_1, u_2)$$

Si $h_1^{-1}(u_1, u_2)$ y $h_2^{-1}(u_1, u_2)$ tienen derivadas parciales continuas con respecto a u_1 y u_2 y jacobiano

$$J = \det \begin{bmatrix} \frac{\partial h_1^{-1}}{\partial u_1} & \frac{\partial h_1^{-1}}{\partial u_2} \\ \frac{\partial h_2^{-1}}{\partial u_1} & \frac{\partial h_2^{-1}}{\partial u_2} \end{bmatrix} = \frac{\partial h_1^{-1}}{\partial u_1} \frac{\partial h_2^{-1}}{\partial u_2} - \frac{\partial h_2^{-1}}{\partial u_1} \frac{\partial h_1^{-1}}{\partial u_2} \neq 0$$

entonces la densidad conjunta de U_1 y U_2 es

$$f_{U_1, U_2}(u_1, u_2) = f_{Y_1, Y_2}(h_1^{-1}(u_1, u_2), h_2^{-1}(u_1, u_2)) |J|$$

donde $|J|$ es el valor absoluto del jacobiano.

Es necesario asegurarse que la transformación bivariente $u_1 = h_1(y_1, y_2)$, $u_2 = h_2(y_1, y_2)$ es una transformación biunívoca para toda (y_1, y_2) tal que $f_{Y_1, Y_2}(y_1, y_2) > 0$. Si esta transformación bivariente no es biunívoca y este método se aplica a ciegas, la función resultante de 'densidad' no tendrá las propiedades necesarias de una función de densidad válida.

Este método se puede extender a más de dos funciones de variables aleatorias. Sea Y_1, Y_2, \dots, Y_k son variables aleatorias continuas conjuntas y

$$U_1 = h_1(Y_1, \dots, Y_k), U_2 = h_2(Y_1, \dots, Y_k), \dots, U_k = h_k(Y_1, \dots, Y_k)$$

donde la transformación

$$u_1 = h_1(y_1, \dots, y_k), u_2 = h_2(y_1, \dots, y_k), \dots, u_k = h_k(y_1, \dots, y_k)$$

es una transformación biunívoca de (y_1, y_2, \dots, y_k) en (u_1, u_2, \dots, u_k) con inversa

$$y_1 = h_1^{-1}(u_1, \dots, u_k), y_2 = h_2^{-1}(u_1, \dots, u_k), \dots, y_k = h_k^{-1}(u_1, \dots, u_k)$$

y $h_1^{-1}(u_1, u_2, \dots, u_k), h_2^{-1}(u_1, u_2, \dots, u_k), \dots, h_k^{-1}(u_1, u_2, \dots, u_k)$ tienen derivadas parciales continuas con respecto a u_1, u_2, \dots, u_k y jacobiano

$$J = \det \begin{bmatrix} \frac{\partial h_1^{-1}}{\partial u_1} & \frac{\partial h_1^{-1}}{\partial u_2} & \dots & \frac{\partial h_1^{-1}}{\partial u_k} \\ \frac{\partial h_2^{-1}}{\partial u_1} & \frac{\partial h_2^{-1}}{\partial u_2} & \dots & \frac{\partial h_2^{-1}}{\partial u_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_k^{-1}}{\partial u_1} & \frac{\partial h_k^{-1}}{\partial u_2} & \dots & \frac{\partial h_k^{-1}}{\partial u_k} \end{bmatrix} \neq 0$$

entonces se puede usar un resultado similar al presentado en esta sección para determinar la densidad conjunta de U_1, U_2, \dots, U_k . Esto requiere calcular el determinante de una matriz $k \times k$.

6.7 Estadísticos de orden

Numerosas funciones de variables aleatorias de interés en la práctica dependen de las magnitudes relativas de las variables observadas. Esto quiere decir que con frecuencia ordenamos variables aleatorias observadas de acuerdo con sus magnitudes. Las variables ordenadas resultantes se denominan *estadísticos de orden*.

Formalmente, denotamos con Y_1, Y_2, \dots, Y_n a variables aleatorias continuas e independientes, con función de distribución $F(y)$ y función de densidad $f(y)$. Denotamos las variables aleatorias ordenadas Y_i por $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$, donde $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$. Usando esta notación,

$$Y_{(1)} = \min(Y_1, Y_2, \dots, Y_n)$$

es la mínima de las variables aleatorias Y_i y

$$Y_{(n)} = \max(Y_1, Y_2, \dots, Y_n)$$

es la máxima de las variables aleatorias Y_i . Las funciones de densidad de probabilidad para $Y_{(1)}$ y $Y_{(n)}$ se pueden determinar usando el método de las funciones de distribución. Primero vamos a deducir la función de densidad de $Y_{(n)}$. Como $Y_{(n)}$ es la máxima de Y_1, Y_2, \dots, Y_n , el evento $(Y_{(n)} \leq y)$ ocurrirá si y sólo si los eventos $(Y_i \leq y)$ ocurren para toda $i = 1, 2, \dots, n$. Esto es

6.6 Transformaciones multivariantes usando jacobianos

Este método se puede utilizar para encontrar la distribución conjunta de varias funciones de variables aleatorias.

Teorema 6.5

Sean Y_1, \dots, Y_n variables aleatorias continuas distribuidas idénticamente e independientes, con función de distribución común $F(y)$ y función de densidad común $f(y)$. Si $Y_{(k)}$ denota el estadístico de orden k -ésimo, entonces la función de densidad de $Y_{(k)}$ está dada por

$$g_{(k)}(y_k) = \frac{n!}{(k-1)!(n-k)!} [F(y_k)]^{k-1} \times [1 - F(y_k)]^{n-k} f(y_k),$$

donde $-\infty < y_k < \infty$. Si j y k son dos enteros tales que $1 \leq j < k \leq n$, la densidad conjunta de $Y_{(j)}$ y $Y_{(k)}$ está dada por

$$g_{(j)(k)}(y_j, y_k) = \frac{n!}{(j-1)!(k-1-j)!(n-k)!} [F(y_j)]^{j-1} \times [F(y_k) - F(y_j)]^{k-1-j} \times [1 - F(y_k)]^{n-k} f(y_j) f(y_k)$$

donde $-\infty < y_j < y_k < \infty$.