

III.- Distribución de función de variables aleatorias (Teoría de la distribución)

3.1 Introducción.

En este capítulo se tratará el estudio de funciones de variables aleatorias, en forma general podemos decir que para X_1, X_2, \dots, X_n variables aleatorias, estaremos interesados en conocer la distribución conjunta de $Y_1 = h_1(X_1, X_2, \dots, X_n), Y_2 = h_2(X_1, X_2, \dots, X_n), \dots, Y_k = h_k(X_1, X_2, \dots, X_n)$. La distribución de Y_1, Y_2, \dots, Y_k , satisface:

$$\begin{aligned} F_{Y_1 Y_2 \dots Y_k}(y_1, y_2, \dots, y_k) &= P(Y_1 \leq y_1, Y_2 \leq y_2, \dots, Y_k \leq y_k) \\ &= P(h_1(X_1, X_2, \dots, X_n) \leq y_1, h_2(X_1, X_2, \dots, X_n) \leq y_2, \dots, h_k(X_1, X_2, \dots, X_n) \leq y_k) \end{aligned}$$

El cálculo de la anterior probabilidad al depender de la X_i 's, se reduciría a integrar o sumar la densidad conjunta sobre la región que determina estas nuevas variables, pero en muchas ocasiones será de mucha utilidad saber explícitamente la distribución conjunta de estas nuevas variables, como lo es en el uso de la inferencia estadística.

3.2 Técnica de la función de distribución acumulativa.

Para X_1, X_2, \dots, X_n variables aleatorias, definimos las siguientes variables de la siguiente forma, $Y_1 = h_1(X_1, X_2, \dots, X_n), Y_2 = h_2(X_1, X_2, \dots, X_n), \dots, Y_k = h_k(X_1, X_2, \dots, X_n)$. La distribución de Y_1, Y_2, \dots, Y_k , satisface:

$$\begin{aligned} F_{Y_1 Y_2 \dots Y_k}(y_1, y_2, \dots, y_k) &= P(Y_1 \leq y_1, Y_2 \leq y_2, \dots, Y_k \leq y_k) \\ &= P(h_1(X_1, X_2, \dots, X_n) \leq y_1, h_2(X_1, X_2, \dots, X_n) \leq y_2, \dots, h_k(X_1, X_2, \dots, X_n) \leq y_k) \end{aligned}$$

Es decir, los eventos $\{ Y_1 \leq y_1, Y_2 \leq y_2, \dots, Y_k \leq y_k \}$ y $\{ b_1(X_1, X_2, \dots, X_n) \leq y_1, b_2(X_1, X_2, \dots, X_n) \leq y_2, \dots, b_k(X_1, X_2, \dots, X_n) \leq y_k \}$ son iguales, si conocemos la distribución conjunta de X_1, X_2, \dots, X_n , podemos calcular la probabilidad del evento $\{ b_1(X_1, X_2, \dots, X_n) \leq y_1, b_2(X_1, X_2, \dots, X_n) \leq y_2, \dots, b_k(X_1, X_2, \dots, X_n) \leq y_k \}$ y por tanto también podemos calcular $F_{Y_1 Y_2 \dots Y_k}(y_1, y_2, \dots, y_k)$. A esta forma de obtener la distribución conjunta de Y_1, Y_2, \dots, Y_k , se le denomina *técnica de la función de distribución acumulativa*.

Veamos unos ejemplos para ver como se aplica dicha técnica.

Ejemplo 3.1. Sea $X \sim N(0,1)$, y $Y = b(X) = X^2$. Apliquemos la técnica de la función de distribución acumulativa, para encontrar la distribución de Y .

$$\begin{aligned}
 F_Y(y) &= P(Y \leq y) \\
 &= P(X^2 \leq y) \\
 &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\
 &= 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \\
 &= \frac{2}{\sqrt{2\pi}} \int_0^y \frac{1}{2u} e^{-\frac{1}{2}\xi} d\xi \\
 &= \int_0^y \frac{1}{\sqrt{2\pi}\sqrt{\xi}} e^{-\xi} d\xi \\
 &= \int_0^y \frac{1}{\sqrt{2\Gamma(\frac{1}{2})}} \xi^{-\frac{1}{2}} e^{-\xi} d\xi
 \end{aligned}$$

La anterior es la distribución de una $\chi_{(1)}^2$.

Ejemplo 3.2. Sea X_1, X_2, \dots, X_n variables aleatorias independientes sea;

$$Y_{(n)} = b(X_1, X_2, \dots, X_n) = \max(X_1, X_2, \dots, X_n).$$

Busquemos la distribución de $Y_{(n)}$, primero veamos que el evento $\{ \max(X_1, X_2, \dots, X_n) \leq y \}$ es equivalente a $\{ X_1 \leq y, X_2 \leq y, \dots, X_n \leq y \}$ luego entonces,

$$\begin{aligned} F_{Y_{(n)}}(y) &= P(Y_{(n)} \leq y) \\ &= P(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y) \end{aligned}$$

Por ser independientes

$$\begin{aligned} &= P(X_1 \leq y)P(X_2 \leq y) \cdots P(X_n \leq y) \\ &= \prod_{i=1}^n F_{X_i}(y) \end{aligned}$$

Si además de independientes asumimos que tienen la misma distribución, tendríamos entonces que

$$F_{Y_{(n)}}(y) = (F_X(y))^n.$$

Para el caso de que las X_i 's sean variables continuas encontramos que

$$f_{Y_{(n)}}(y) = n(F_X(y))^{n-1} f_X(y)$$

Ejemplo 3.3. Ahora busquemos la distribución de $Y_{(1)} = \min(X_1, X_2, \dots, X_n)$, los eventos $\{\min(X_1, X_2, \dots, X_n) > y\}$ es equivalente a $\{X_1 > y, X_2 > y, \dots, X_n > y\}$ luego

$$\begin{aligned} F_{Y_{(1)}}(y) &= 1 - P(Y_{(1)} > y) \\ &= 1 - P(X_1 > y, X_2 > y, \dots, X_n > y) \end{aligned}$$

Por ser independientes

$$\begin{aligned} &= 1 - [(1 - P(X_1 \leq y))(1 - P(X_2 \leq y)) \cdots (1 - P(X_n \leq y))] \\ &= 1 - \prod_{i=1}^n (1 - F_{X_i}(y)) \end{aligned}$$

Si además de independientes asumimos que tienen la misma distribución, tendríamos entonces que

$$F_{Y_{(1)}}(y) = 1 - (1 - F_X(y))^n.$$

Para el caso de que las X_i 's sean variables continuas encontramos que

$$f_{Y_{(1)}}(y) = n(1 - F_X(y))^{n-1} f_X(y)$$

Ejemplo 3.4. Ahora sean X y Y dos variables aleatorias continuas distribuidas conjuntamente, busquemos la distribución de $Z = X + Y$.

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X+Y \leq z) \\ &= \iint_{x+y \leq z} f_{XY}(x, y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{XY}(x, y) dy dx \end{aligned}$$

Haciendo $u = y + x$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^z f_{XY}(x, u-x) du dx$$

Ahora derivando con respecto a z

$$\begin{aligned} f_Z(z) &= \frac{dF_Z(z)}{dz} \\ &= \frac{d}{dz} \int_{-\infty}^z \int_{-\infty}^z f_{XY}(x, u-x) du dx \\ &= \int_{-\infty}^{\infty} f_{XY}(x, z-x) dx \end{aligned}$$

De forma similar podemos encontrar que para $V = X - Y$

$$f_V(v) = \int_{-\infty}^{\infty} f_{XY}(x, x-v) dx$$

Si ambas variables son independientes lo anterior nos queda como:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

y

$$f_V(v) = \int_{-\infty}^{\infty} f_X(x) f_Y(x-v) dx$$

Ejemplo 3.5. Ahora sean X y Y dos variables aleatorias continuas distribuidas conjuntamente, busquemos la distribución de $Z = XY$.

$$\begin{aligned} F_Z(\tilde{z}) &= P(Z \leq \tilde{z}) = P(XY \leq \tilde{z}) \\ &= \iint_{xy \leq \tilde{z}} f_{XY}(x, y) dy dx \\ &= \int_{-\infty}^0 \int_{\tilde{z}/x}^{\infty} f_{XY}(x, y) dy dx + \int_0^{\tilde{z}/x} \int_{-\infty}^{\infty} f_{XY}(x, y) dy dx \end{aligned}$$

Haciendo $u = yx$

$$\begin{aligned} &= \int_{-\infty}^0 \int_{\tilde{z}}^{\infty} f_{XY}\left(x, \frac{u}{x}\right) \frac{du}{x} dx + \int_0^{\tilde{z}} \int_{-\infty}^{\infty} f_{XY}\left(x, \frac{u}{x}\right) \frac{du}{x} dx \\ &= \int_{-\infty}^0 \int_{\tilde{z}}^{\infty} f_{XY}\left(x, \frac{u}{x}\right) \frac{du}{x} dx + \int_0^{\tilde{z}} \int_{-\infty}^{\infty} f_{XY}\left(x, \frac{u}{x}\right) \frac{du}{x} dx \\ &= \int_{-\infty}^0 \int_{-\infty}^{\tilde{z}} f_{XY}\left(x, \frac{u}{x}\right) \frac{du}{-x} dx + \int_0^{\tilde{z}} \int_{-\infty}^{\infty} f_{XY}\left(x, \frac{u}{x}\right) \frac{du}{x} dx \\ &= \int_{-\infty}^{\tilde{z}} \int_{-\infty}^{\infty} \frac{1}{|x|} f_{XY}\left(x, \frac{u}{x}\right) du dx \end{aligned}$$

Entonces

$$\begin{aligned} f_Z(\tilde{z}) &= \frac{dF_Z(\tilde{z})}{d\tilde{z}} \\ &= \frac{d}{d\tilde{z}} \int_{-\infty}^{\tilde{z}} \int_{-\infty}^{\infty} \frac{1}{|x|} f_{XY}\left(x, \frac{u}{x}\right) dx du \\ &= \int_{-\infty}^{\infty} \frac{1}{|x|} f_{XY}\left(x, \frac{\tilde{z}}{x}\right) dx \end{aligned}$$

De forma similar podemos encontrar que para $V = \frac{X}{Y}$

$$f_V(v) = \int_{-\infty}^{\infty} |x| f_{XY}(x, vx) dx$$

3.3 Técnica de la función generadora de momentos.

Para X_1, X_2, \dots, X_n variables aleatorias, con función de densidad conjunta $f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$ definimos las siguientes variables de la siguiente forma, $Y_1 = b_1(X_1, X_2, \dots, X_n)$, $Y_2 = b_2(X_1, X_2, \dots, X_n)$, ..., $Y_k = b_k(X_1, X_2, \dots, X_n)$. La función generadora de Y_1, Y_2, \dots, Y_k , si existe esta es:

$$\begin{aligned} m_{Y_1 Y_2 \dots Y_k}(t_1, t_2, \dots, t_k) &= E\left(e^{\langle t, Y \rangle}\right) = E\left(e^{Y_1 t_1 + Y_2 t_2 + \dots + Y_k t_k}\right) \\ &= E\left(e^{b_1(X_1, X_2, \dots, X_n) t_1 + b_2(X_1, X_2, \dots, X_n) t_2 + \dots + b_k(X_1, X_2, \dots, X_n) t_k}\right) \end{aligned}$$

Como se puede ver la generadora de momentos es una función de las variables X_1, X_2, \dots, X_n , así que esta la podemos calcular ya que conocemos su función de densidad conjunta, una vez que se obtiene la generadora de momentos, después hay que ver a que distribución esta asociada, las limitantes de estas técnicas es que se conoce la generadora de momentos de varias variables aleatorias para el caso univariado pero para el caso multivariado son pocas las generadoras conocidas.

Ejemplo 3.6. Sea $X \sim N(0,1)$, y $Y = b(X) = X^2$. Apliquemos la técnica de la función generadora de momentos, para encontrar la distribución de Y .

$$\begin{aligned} m_Y(t) &= E(e^{tY}) \\ &= E\left(e^{tX^2}\right) \\ &= \int_{-\infty}^{\infty} e^{tx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2(1-2t)} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \frac{du}{(1-2t)^{\frac{1}{2}}} \\ &= \frac{1}{(1-2t)^{\frac{1}{2}}} \end{aligned}$$

La anterior es la generadora de momentos de una $\chi_{(1)}^2$, así X^2 tiene distribución $\chi_{(1)}^2$.

Ejemplo 3.7. Sea X_1 y X_2 dos variables aleatorias independientes con distribución $N(0,1)$, sea $Y_1 = X_1 - X_2$ $Y_2 = X_1 + X_2$. Apliquemos la técnica de la función generadora de momentos, para encontrar la distribución conjunta de Y_1 y Y_2 .

$$\begin{aligned} m_{Y_1 Y_2}(t_1, t_2) &= E\left(e^{Y_1 t_1 + Y_2 t_2}\right) \\ &= E\left(e^{(X_1 - X_2)t_1 + (X_1 + X_2)t_2}\right) \\ &= E\left(e^{X_1(t_1 + t_2) + X_2(t_2 - t_1)}\right) \end{aligned}$$

Como X_1 y X_2 son independientes

$$\begin{aligned} &= E\left(e^{X_1(t_1 + t_2)}\right) E\left(e^{X_2(t_2 - t_1)}\right) \\ &= e^{\frac{(t_1 + t_2)^2}{2}} e^{\frac{(t_2 - t_1)^2}{2}} \\ &= e^{t_1^2 + t_2^2} \\ &= e^{\frac{2t_1^2}{2}} e^{\frac{2t_2^2}{2}} \\ &= m_{Y_1}(t_1) m_{Y_2}(t_2) \end{aligned}$$

Tenemos que Y_1 y Y_2 cada una se distribuye como una $N(0,2)$ y ambas son independientes.

Nota. Recordemos que para $Y = aX + b$, con a, b constantes, tenemos que la generadora de momentos de Y es: $m_Y(t) = e^{tb} m_X(at)$

Si X_1, X_2, \dots, X_n son variables aleatorias independientes, definamos a $S = \sum_{i=1}^n X_i$ la generadora de momentos de S es:

$$\begin{aligned} m_S(t) &= E\left(e^{tS}\right) \\ &= E\left(e^{tX_1 + tX_2 + \dots + tX_n}\right) \\ &= \prod_{i=1}^n m_{X_i}(t) \end{aligned}$$

Si además de independencia la X_i 's son idénticamente distribuidas tenemos que

$$m_S(t) = \left(m_X(t)\right)^n$$

Ejemplo 3.8. Sea X_1, X_2, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas (i.i.d), con distribución $\text{Poi}(\lambda)$, busquemos la distribución de $S = \sum_{i=1}^n X_i$.

$$m_S(t) = (m_X(t))^n = \left(e^{\lambda(e^t-1)} \right)^n = e^{n\lambda(e^t-1)}$$

está es la distribución de una $\text{Poi}(n\lambda)$.

Ejemplo 3.9. Sea X_1, X_2, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas (i.i.d), con distribución $\text{Exp}(\lambda)$, busquemos la distribución de $S = \sum_{i=1}^n X_i$.

$$m_S(t) = (m_X(t))^n = \left(\frac{\lambda}{\lambda + t} \right)^n$$

está es la distribución de una $\text{Gam}(n, \lambda)$.

3.4 Transformaciones.

Primero empezaremos viendo el uso de esta técnica para el caso de una sola variable aleatoria, para después extenderlo, al buscar la densidad de una función de la variable aleatoria lo que estamos haciendo es una transformación, que es determinada por $Y=h(X)$.

Si X es una variable aleatoria discreta, si X toma a x_1, x_2, \dots, x_n , como sus posibles valores con probabilidad $f_X(x_1), f_X(x_2), \dots, f_X(x_n)$, los posibles valores de Y se obtienen de sustituir los valores de X en $h(\cdot)$, así tendemos que varios valores de X pueden dar el mismo valor de Y , así la probabilidad de que salga el valor y_j de Y sería la suma de las probabilidades de los valores de X que dan este valor, esto es:

$$f_Y(y_j) = \sum_{\{i: h(x_i)=y_j\}} f_X(x_i)$$

Ejemplo 3.10. Supongamos que X es una v.a con valores $-2, 0, 2$ con probabilidad

$$f_X(-2) = p_1, f_X(0) = p_2, f_X(2) = p_3, \text{ con } p_1 + p_2 + p_3 = 1 \text{ y } p_i > 0,$$

sea $Y = X^2 + 1$, es claro que Y es una variable discreta que toma los valores de 1 y de 5 y su función de probabilidad (función de masa de probabilidad)

$$f_Y(1) = P_X\{x: x^2+1 = 1\} = P_X\{x: x^2 = 0\} = f_X(0) = p_2$$

$$f_Y(5) = P_X\{x: x^2+1 = 5\} = P_X\{x: x = \pm 2\} = f_X(-2) + f_X(2) = p_1 + p_3$$

Ahora si X es una variable aleatoria continua, con densidad $f_X(\cdot)$, sea $D_X = \{x: f_X(\cdot) > 0\}$, si a D_X lo podemos descomponer en un número finito de conjuntos disjuntos $D_{X_1}, D_{X_2}, \dots, D_{X_m}$, tal que $y = b(x)$ defina una transformación uno a uno de cada uno de los conjuntos $D_{X_1}, D_{X_2}, \dots, D_{X_m}$ en D_Y con $y \in D_Y$, entonces la densidad conjunta de $Y = b(X)$ se puede obtener. Sea $x = b_i^{-1}(y)$ denota la inversa de $y = b(x)$ para x en D_{X_i} , y la derivada de $x = b_i^{-1}(y)$ con respecto a y es continua y no cero para $y \in D_Y$, entonces la densidad de $Y = b(X)$ está dada por:

$$f_Y(y) = \sum_{i=1}^m \left| \frac{d}{dy} b_i^{-1}(y) \right| f_X(b_i^{-1}(y)) I_{D_Y}(y)$$

Ejemplo 3.11. Supongamos que X es una v.a continua con distribución $U(0,1)$, busquemos la densidad de $Y = -\ln X$. La densidad de X es

$$f_X(x) = I_{(0,1)}(x)$$

y

$$b^{-1}(y) = e^{-y}$$

Así tenemos que

$$f_Y(y) = \left| \frac{d}{dy} e^{-y} \right| f_X(e^{-y}) = e^{-y} I_{(0,1)}(e^{-y}) = e^{-y} I_{(0,\infty)}(y)$$

Podemos ver que Y se distribuye como una $\text{Exp}(1)$.

Ejemplo 3.12. Sea $X \sim N(0,1)$, y $Y = b(X) = X^2$. Tenemos que por ser una normal X toma valores en toda la recta real y nuestra función deja de ser uno a uno en toda la

recta pero si dividimos a D_X en $D_{X_1} \{x : x \in D_X, x < 0\}$ y $D_{X_2} \{x : x \in D_X, x \geq 0\}$, obtenemos $b_1^{-1}(y) = -\sqrt{y}$ y $b_2^{-1}(y) = \sqrt{y}$, son transformaciones uno a uno en las regiones que definimos, así tenemos que la densidad de Y es:

$$\begin{aligned} f_Y(y) &= \left| \frac{d}{dy} -\sqrt{y} \right| f_X(-\sqrt{y}) + \left| \frac{d}{dy} \sqrt{y} \right| f_X(\sqrt{y}) \\ &= \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) \end{aligned}$$

Por ser X una normal

$$\begin{aligned} &= \frac{1}{\sqrt{y}} f_X(\sqrt{y}) \\ &= \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} I_{(0,\infty)}(\sqrt{y}) \\ &= \frac{1}{\sqrt{2}\Gamma(\frac{1}{2})} y^{-\frac{1}{2}} e^{-y} \end{aligned}$$

Y como ya lo habíamos encontrado anteriormente Y es la distribución de una $\chi_{(1)}^2$.

Ejemplo 3.13. Sea X una v.a continua, con densidad

$$f_X(x) = \left(\frac{x}{3}\right)^2 I_{[0,3]}(x)$$

sea $Y = \begin{cases} -2 & \text{si } X < 2 \\ 0 & \text{si } X = 2 \\ 2 & \text{si } X > 2 \end{cases}$, la densidad de Y sería

$$f_Y(2) = \int_0^2 \left(\frac{x}{3}\right)^2 I_{[0,3]}(x) dx = \frac{8}{27}$$

$$f_Y(0) = 0$$

$$f_Y(2) = \int_2^3 \left(\frac{x}{3}\right)^2 I_{[0,3]}(x) dx = \frac{19}{27}$$

A continuación enunciaremos un resultado muy importante, para la simulación de variables aleatorias.

Teorema 3.1. Si X es una v.a continua con función de distribución $F_X(x)$, entonces $U = F_X(x)$ se distribuye como $U(0,1)$, Inversamente si U se distribuye como $U(0,1)$, entonces $X = F_X^{-1}(U)$ tiene a $F_X(x)$ como función de distribución acumulativa.

Demostración.

\Rightarrow

$$P(U \leq u) = P(F_X(x) \leq u) = P(X \leq F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u.$$

Esta es la distribución de una $U(0,1)$.

\Leftarrow

$$P(X \leq x) = P(F_X^{-1}(U) \leq x) = P(U \leq F_X(x)) = F_X(x)$$

□

Nota. Al resultado anterior se le suele llamar como la *transformación integral de la probabilidad*.

Ahora extendamos el tema de las transformaciones para más de una dimensión, empecemos con las discretas.

Sea $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ la densidad conjunta del vector aleatorio discreto $\mathbf{X} = (X_1, X_2, \dots, X_n)$, Sea $D_{\mathbf{X}} = \{(x_1, x_2, \dots, x_n) : f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) > 0\}$, estamos interesados en conocer la densidad conjunta de $Y_1 = h_1(X_1, X_2, \dots, X_n)$, $Y_2 = h_2(X_1, X_2, \dots, X_n)$, ..., $Y_k = h_k(X_1, X_2, \dots, X_n)$. La densidad la podemos encontrar de la siguiente manera:

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_k = y_k) = f_{Y_1, Y_2, \dots, Y_k}(y_1, y_2, \dots, y_k) = \sum f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$$

Donde la suma es sobre todos los posibles valores (x_1, x_2, \dots, x_n) que pertenecen a $D_{\mathbf{X}}$ para los cuales $(y_1, y_2, \dots, y_k) = (h_1((x_1, x_2, \dots, x_n)), h_2((x_1, x_2, \dots, x_n)), \dots, h_k((x_1, x_2, \dots, x_n)))$.

Ejemplo 3.14. Sea (X_1, X_2, X_3) con la siguiente función de densidad,

(x_1, x_2, x_3)	$(0,0,0)$	$(0,0,1)$	$(0,1,1)$	$(1,0,1)$	$(1,1,0)$	$(1,1,1)$
$f_{X_1X_2X_3}(x_1, x_2, x_3)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Obtengamos la densidad conjunta de $Y_1 = X_1 + X_2 + X_3$ y $Y_2 = |X_3 - X_2|$

$$\begin{aligned}
 f_{Y_1Y_2}(0,0) &= f_{X_1X_2X_3}(0,0,0) = \frac{1}{8} \\
 f_{Y_1Y_2}(1,1) &= f_{X_1X_2X_3}(0,0,1) = \frac{3}{8} \\
 f_{Y_1Y_2}(2,0) &= f_{X_1X_2X_3}(0,1,1) = \frac{1}{8} \\
 f_{Y_1Y_2}(2,1) &= f_{X_1X_2X_3}(1,0,1) + f_{X_1X_2X_3}(1,1,0) = \frac{2}{8} \\
 f_{Y_1Y_2}(3,0) &= f_{X_1X_2X_3}(1,1,1) = \frac{1}{8}
 \end{aligned}$$

Para el caso continuo tenemos algo muy similar al caso univariado, primero definamos un resultado que nos ayudará a encontrar la densidad conjunta de nuestras nuevas variables.

Si $\mathbf{X} = (X_1, X_2, \dots, X_n)$, es un vector de v.a continuas, sea $D_{\mathbf{X}} = \{(x_1, x_2, \dots, x_n) : f_{X_1X_2\dots X_n}(x_1, x_2, \dots, x_n) > 0\}$, si a $D_{\mathbf{X}}$ lo podemos descomponer en un número finito de conjuntos disjuntos $D_{X_1}, D_{X_2}, \dots, D_{X_m}$, tal que $y_1 = h_1(x_1, x_2, \dots, x_n)$, $y_2 = h_2(x_1, x_2, \dots, x_n)$, $\dots, y_n = h_n(x_1, x_2, \dots, x_n)$ son transformaciones uno a uno de D_{X_i} dentro de $D_{\mathbf{Y}}$, $\mathbf{Y} \in D_{\mathbf{Y}}$, con $i = 1, 2, \dots, m$. Sea $x_1 = h_{1i}^{-1}(y_1, y_2, \dots, y_n)$, \dots , $x_n = h_{ni}^{-1}(y_1, y_2, \dots, y_n)$ la transformación inversa $D_{\mathbf{Y}}$ dentro de cada D_{X_i} con $i = 1, 2, \dots, m$. Sea

$$J_i = \begin{vmatrix} \frac{dh_{1i}^{-1}}{dy_1} & \frac{dh_{1i}^{-1}}{dy_2} & \dots & \frac{dh_{1i}^{-1}}{dy_n} \\ \frac{dh_{2i}^{-1}}{dy_1} & \frac{dh_{2i}^{-1}}{dy_2} & \dots & \frac{dh_{2i}^{-1}}{dy_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dh_{ni}^{-1}}{dy_1} & \frac{dh_{ni}^{-1}}{dy_2} & \dots & \frac{dh_{ni}^{-1}}{dy_n} \end{vmatrix}$$

para $i = 1, 2, \dots, m$. Asumimos que todas las derivadas parciales son continuas sobre $D_{\mathbf{Y}}$ y el determinante J_i es distinto de cero, para todo i , entonces

$$f_{Y_1 Y_2 \dots Y_k}(y_1, y_2, \dots, y_k) = \sum_{i=1}^m |J_i| f_{X_1 X_2 \dots X_n}(b_{1i}^{-1}(y_1, y_2, \dots, y_n), b_{2i}^{-1}(y_1, y_2, \dots, y_n), \dots, b_{ni}^{-1}(y_1, y_2, \dots, y_n))$$

para $(y_1, y_2, \dots, y_n) \in D_Y$.

Ejemplo 3.15. Sea X_1 y X_2 dos variables aleatorias independientes con distribución $N(0,1)$. Sea $Y_1 = X_1 + X_2$ y $Y_2 = X_1/X_2$. Entonces

$$x_1 = b_1^{-1}(y_1, y_2) = \frac{y_1 y_2}{1 + y_2} \quad \text{y} \quad x_2 = b_2^{-1}(y_1, y_2) = \frac{y_1}{1 + y_2}$$

$$J = \begin{vmatrix} \frac{y_2}{1 + y_2} & \frac{y_1}{(1 + y_2)^2} \\ \frac{1}{1 + y_2} & -\frac{y_1}{(1 + y_2)^2} \end{vmatrix} = -\frac{y_1(1 + y_2)}{(1 + y_2)^3} = -\frac{y_1}{(1 + y_2)^2}$$

$$f_{Y_1 Y_2}(y_1, y_2) = \frac{|y_1|}{(1 + y_2)^2} \frac{1}{2\pi} e^{-\frac{1}{2} \left[\frac{(y_1 y_2)^2}{(1 + y_2)^2} + \frac{y_1^2}{(1 + y_2)^2} \right]}$$

De la suma de dos normales sabemos que es un normal, pero veamos cual es la distribución de Y_2 obteniendo la marginal de la conjunta

$$\begin{aligned} f_{Y_2}(y_2) &= \int_{-\infty}^{\infty} \frac{|y_1|}{(1 + y_2)^2} \frac{1}{2\pi} e^{-\frac{1}{2} \left[\frac{(y_1 y_2)^2}{(1 + y_2)^2} + \frac{y_1^2}{(1 + y_2)^2} \right]} dy_1 \\ &= \frac{1}{2\pi(1 + y_2)^2} \int_{-\infty}^{\infty} |y_1| e^{-\frac{1}{2} \left[\frac{(1 + y_2^2) y_1^2}{(1 + y_2)^2} \right]} dy_1 \end{aligned}$$

sea

$$u = \frac{1}{2} \frac{(1 + y_2^2) y_1^2}{(1 + y_2)^2}$$

entonces

$$du = \frac{1}{2} \frac{(1 + y_2^2)}{(1 + y_2)^2} y_1 dy_1$$

y así

$$f_{Y_2}(y_2) = \frac{1}{2\pi(1+y_2)^2} \frac{(1+y_2)^2}{1+y_2^2} 2 \int_0^\infty e^{-u} du = \frac{1}{\pi(1+y_2^2)}$$

Que es la densidad de una Cauchy.

Ejemplo 3.16. Sea X_1 y X_2 dos variables aleatorias independientes con distribución $\text{Gam}(n_i, \lambda)$ con $i = 1, 2$, respectivamente. Supongamos que deseamos obtener la distribución de $Y_1 = \frac{X_1}{X_1 + X_2}$. Aquí solamente tenemos una función, lo que tenemos

que hacer es proponer otra, en el exponente tenemos que este depende de la suma de las dos variables, además de que nos ayuda el saber que las suma de gammas nos da otra gamma así que parece ser que $Y_2 = X_1 + X_2$ es una buena elección, luego entonces tenemos que

$$x_1 = h_1^{-1}(y_1, y_2) = y_1 y_2 \text{ y } x_2 = h_2^{-1}(y_1, y_2) = y_2 - y_1 y_2$$

y

$$J = \begin{vmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{vmatrix} = y_2$$

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= y_2 \frac{1}{\Gamma(n_1)} \frac{1}{\Gamma(n_2)} \lambda^{n_1+n_2} (y_1 y_2)^{n_1-1} (y_2 - y_1 y_2)^{n_2-1} e^{-\lambda y_2} I_{(0, \infty)}(y_1 y_2) I_{(0, \infty)}(y_2 - y_1 y_2) \\ &= \frac{\lambda^{n_1+n_2}}{\Gamma(n_1) \Gamma(n_2)} y_1^{n_1-1} (1 - y_1)^{n_2-1} y_2^{n_1+n_2-1} e^{-\lambda y_2} I_{(0, 1)}(y_1) I_{(0, \infty)}(y_2) \end{aligned}$$

Tenemos que $B(n_1, n_2) = \frac{\Gamma(n_1) \Gamma(n_2)}{\Gamma(n_1 + n_2)}$

$$f_{Y_1, Y_2}(y_1, y_2) = \left[\frac{1}{B(n_1, n_2)} y_1^{n_1-1} (1 - y_1)^{n_2-1} I_{(0, 1)}(y_1) \right] \left[\frac{\lambda^{n_1+n_2}}{\Gamma(n_1 + n_2)} y_2^{n_1+n_2-1} e^{-\lambda y_2} I_{(0, \infty)}(y_2) \right]$$

De aquí tenemos que Y_1 y Y_2 son independientes y que Y_1 , se distribuye como una beta.

3.5 Normal Multivariada

Definición 3.1. Si $X^tAX > 0$ se dice que X^tAX es definida positiva, para todo $X \neq 0$ y A es positiva.

Definición 3.2. Si $X^tAX \geq 0$ se dice que X^tAX es semipositiva, para todo $X \neq 0$ y A es semipositiva.

Definición 3.3. Si A es positiva o semipositiva A es no negativa.

Lema 3.1. La matriz A es positiva si, y sólo si todos los determinantes de sus matrices angulares son positivos.

Lema 3.2. Si P es no singular P^tAP es o no es definida positiva (semi) si A es o no lo es.

Lema 3.3. Matrices definidas positivas son no singulares.

Lema 3.4. Matrices semipositivas son singulares (el caso contrario no siempre ocurre es decir una matriz singular no es siempre semipositiva).

Lema 3.5. Las raíces de una matriz positiva (semi) son todas mayores que cero (mayor o igual a cero).

Lema 3.6. AA^t es positiva cuando A es de rango completo por filas y semipositiva en cualquier otro caso.

Lema 3.7. A^tA es positiva cuando A tiene rango completo por columna y semipositiva en cualquier otro caso.

Lema 3.8. Si A es simétrica de orden n y rango r puede escribirse como $A = LL^t$, L es de $n \times r$ de rango r .

Lema 3.9. Si A es simétrica, es positiva si, y sólo si puede escribirse como PP^t con P no singular.

Teorema 3.2. Si X es un vector de $p \times 1$ con elementos x_i y sea A un vector de $p \times 1$ con

elementos a_i y sea $Z = X^t A = A^t X$ la derivada de Z con respecto a X es $\frac{\partial Z}{\partial X} = \begin{pmatrix} \frac{\partial Z}{\partial x_1} \\ \vdots \\ \frac{\partial Z}{\partial x_p} \end{pmatrix} = A$

Demostración.

El i -ésimo elemento de $\frac{\partial Z}{\partial X}$ es;

$$\frac{\partial Z}{\partial x_i} = \frac{\partial \left(\sum_{j=1}^p a_j x_j \right)}{\partial x_i}$$

así que el i -ésimo elemento de $\frac{\partial Z}{\partial X}$ es a_i , por lo que $\frac{\partial Z}{\partial X} = A$

□

Teorema 3.3. Sea A un vector de $p \times 1$ y B de $q \times 1$ y X una matriz de $p \times q$ cuyo ij -ésimo

elemento es x_{ij} . Sea $Z = A^t X B = \sum_{m=1}^q \sum_{n=1}^p a_n x_{nm} b_m$, entonces $\frac{\partial Z}{\partial X} = AB^t$.

Demostración.

El ij -ésimo elemento de $\frac{\partial Z}{\partial X}$ es

$$\frac{\partial Z}{\partial x_{ij}} = \frac{\partial \left(\sum_{m=1}^q \sum_{n=1}^p a_n x_{nm} b_m \right)}{\partial x_{ij}} = a_i b_j$$

por tanto se sigue que $\frac{\partial Z}{\partial X} = AB^t$.

□

Teorema 3.4. Sea A una matriz simétrica de $p \times p$ y X de $p \times 1$. Sea $Z = X^t A X$, entonces

$\frac{\partial Z}{\partial A} = 2XX^t - D(XX^t)$, donde $D(XX^t)$ es la matriz diagonal cuyos elementos son la diagonal de XX^t .

Demostración.

El ij -ésimo elemento de $\frac{\partial Z}{\partial A}$ es

$$\frac{\partial Z}{\partial a_{ij}} = \frac{\partial \left(\sum_{m=1}^p \sum_{n=1}^p x_n x_m a_{mn} \right)}{\partial a_{ij}}$$

si $i = j$, $\frac{\partial Z}{\partial a_{ij}} = x_i^2$. Si $i \neq j$, $\frac{\partial Z}{\partial a_{ij}} = 2x_i x_j$. Así tenemos que $\frac{\partial Z}{\partial A} = 2XX^t - D(XX^t)$.

□

Teorema 3.5. Sea A una matriz simétrica de $p \times p$ y X de $p \times 1$. Sea $Z = X^t A X$, entonces

$$\frac{\partial Z}{\partial X} = 2AX.$$

Demostración.

El ij -ésimo elemento de $\frac{\partial Z}{\partial X}$ es

$$\begin{aligned} \frac{\partial Z}{\partial x_i} &= \frac{\partial \left(\sum_{m=1}^p \sum_{n=1}^p x_n x_m a_{mn} \right)}{\partial x_i} \\ &= \frac{\partial \left(\sum_{m=1}^p x_m^2 a_{mm} + \sum_{\substack{m=1 \\ m \neq n}}^p \sum_{n=1}^p x_n x_m a_{mn} \right)}{\partial x_i} \\ &= 2x_i a_{ii} + 2 \sum_{\substack{n=1 \\ n \neq i}}^p x_n a_{in} = 2 \sum_{n=1}^p x_n a_{in} \end{aligned}$$

luego entonces $\frac{\partial Z}{\partial X} = 2AX$.

□

La generalización de la función de densidad de una distribución Normal a varias dimensiones juega un papel importante en el análisis de regresión.

Una ventaja de la distribución normal multivariada parte del hecho de que es matemáticamente tratable y se pueden obtener resultados “buenos”. Frecuentemente este no es el caso con otras distribuciones multivariadas.

Antes de definir la distribución Normal Multivariada, necesitamos definir algunos momentos de un vector aleatorio, es decir, un vector cuyos componentes están

distribuidos conjuntamente. La media o esperanza de un vector aleatorio $X = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$ de

$m \times 1$ esta definido por ser el vector de esperanzas:

$$E(X) = \begin{pmatrix} E(x_1) \\ \vdots \\ E(x_m) \end{pmatrix}$$

Mas generalmente, si $Z = (z_{ij})$ es una matriz aleatoria de $p \times q$, entonces $E(Z)$, la esperanza de Z , es la matriz cuyo ij -ésimo elemento es $E(z_{ij})$. Aunque no entremos en detalle del desarrollo del siguiente resultado, es simple comprobar que si B , C y D son matrices de constantes de $m \times p$, $q \times n$ y $m \times n$ respectivamente, entonces

$$E(BZC + D) = B E(Z) C + D \quad (1.1)$$

Si X tiene media μ la *matriz de varianza-covarianza (varcov)* de X esta definida por la matriz de

$$\Sigma = \text{Cov}(X) = E[(X - \mu)(X - \mu)^t].$$

El elemento ij -ésimo de Σ es

$$\sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)],$$

la covarianza entre las variables x_i y x_j , y el ii -ésimo elemento es

$$\sigma_{ii} = E[(x_i - \mu_i)^2],$$

la varianza de x_i , la matriz de varianza-covarianza la podemos poner como

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

es decir, es una matriz de orden p y donde σ_{ji} es la covarianza entre la variable j e i , como podemos ver los elementos de la diagonal de Σ son no negativos. Σ es simétrica,

es decir, $\Sigma = \Sigma^t$. Además, la clase de matrices de varianza-covarianza coincide con la clase de matrices no-negativas definidas. Recordemos que una matriz simétrica A de $m \times m$ es no negativa definida si

$$\alpha^t A \alpha \geq 0 \quad \text{para toda } \alpha \in \mathbb{R}^m$$

y definida positiva si

$$\alpha^t A \alpha > 0 \quad \text{para toda } \alpha \in \mathbb{R}^m.$$

Teorema 3.6. La matriz Σ de $m \times m$ es una matriz de varianza-covarianza, si, y solo si es no-negativa definida.

Demostración.

Supongamos que Σ es la matriz de varianza-covarianza de un vector aleatorio X , donde X tiene media μ , entonces para toda $\alpha \in \mathbb{R}^{m \times 1}$,

$$\begin{aligned} \text{Var}(\alpha^t X) &= E[(\alpha^t X - \alpha^t \mu)(\alpha^t X - \alpha^t \mu)^t] & (1.2) \\ &= E[\alpha^t (X - \mu)(X - \mu)^t \alpha] \\ &= \alpha^t \Sigma \alpha \geq 0 \end{aligned}$$

luego Σ es no negativa definida.

Ahora supongamos que Σ es una matriz no-negativa definida de rango r , digamos ($r \leq m$). Escribimos $\Sigma = C C^t$, donde C es una matriz de $m \times r$ de rango r . Sea Y un vector de $r \times 1$ de variables aleatorias independientes con media 0 y $\text{Cov}(Y) = I$ para $X = CY$. Entonces $E(X) = 0$ y

$$\begin{aligned} \text{Cov}(X) &= E[XX^t] = E[CY Y^t C^t] \\ &= CE(Y Y^t)C^t \\ &= CC^t = \Sigma, \end{aligned}$$

tal que Σ es una matriz varcov. □

Comúnmente hacemos transformaciones lineales de vectores aleatorios y necesitamos saber como las matrices de varianza-covarianza son transformadas. Supongamos que X es un vector aleatorio de $m \times 1$ con media μ_x y matriz de varianza-covarianza Σ_x y sea $Y = BX + b$, donde B es de $k \times m$ y b es de $k \times 1$. la media de Y es, $\mu_y = B \mu_x + b$, y la matriz de varianza-covarianza de Y es;

$$\begin{aligned} \Sigma_y &= E[(Y - \mu_y)(Y - \mu_y)^t] & (1.3) \\ &= E[(BX + b - (B\mu_x + b))(BX + b - (B\mu_x + b))^t] \end{aligned}$$

$$\begin{aligned}
 &= \text{BE}[(X - \mu_X) (X - \mu_X)^t] B^t \\
 &= B \Sigma B^t
 \end{aligned}$$

Recordemos que la distribución normal univariada, con media μ y varianza σ^2 , tiene función de densidad:

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

ahora considerando el exponente:

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu) (\sigma^2)^{-1} (x-\mu)$$

podemos generalizarlo para un vector de las observaciones sobre varias variables, X de $p \times 1$ tal que tendríamos

$$(X - \mu_X)^t \Sigma^{-1} (X - \mu_X)$$

donde el vector μ_X de $p \times 1$ representa el valor esperado del vector aleatorio X y la matriz Σ es la matriz de varianza covarianza.

La constante de la normal univariada $\frac{1}{\sqrt{2\pi\sigma^2}}$ se sustituya por una constante mas general, la constante para la distribución normal multivariada es $(2\pi)^{-p/2} |\Sigma|^{-1/2}$, consecuentemente la *densidad normal p -dimensional* para el vector aleatorio X tiene la forma:

$$f_X(X) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(X-\mu_X)^t \Sigma^{-1} (X-\mu_X)}$$

donde $-\infty < x_i < \infty$, $i = 1, \dots, p$. y Σ es de rango p . La media de X esta definida por $E[X] = \mu$, y la matriz de varianza covarianza por $E[(X - \mu_X)(X - \mu_X)^t] = \Sigma$.

Usualmente la densidad normal p -dimensional se denota por $N_p(\mu, \Sigma)$, la cual es análoga al caso univariado.

Ahora daremos la definición formal de la densidad normal multivariada.

Definición 3.4. Si y_1, y_2, \dots, y_p son p variables aleatorias y si Y es el vector $p \times 1$ de estas variables

$$f(y_1, y_2, \dots, y_p) = K e^{-\frac{1}{2}(Y-\mu)^t R (Y-\mu)} \quad -\infty < y_i < \infty, i = 1, \dots, p$$

es la *función de densidad de una normal multivariada* si se cumple

- a) R es una matriz definida positiva, r_{ij} son constantes.
- b) K es una constante positiva.
- c) μ_i es el i -ésimo elemento de μ , μ_i son constantes.

$$K = \frac{|R|^{1/2}}{(2\pi)^{p/2}}$$

Haciendo $R = \Sigma^{-1}$, tenemos la misma función deducida anteriormente a partir de la función de densidad normal univariada.

Otra definición alterna es la siguiente.

Definición 3.5. El vector aleatorio de $m \times 1$ X se dice que tiene una distribución normal m -variada si, para cada $\alpha \in \mathbb{R}^{m \times 1}$, la distribución de $\alpha^t X$ es normal univariada.

A partir de las definiciones ahora estableceremos algunas propiedades de la distribución normal multivariada.

Teorema 3.7. Si X es $N_m(\mu, \Sigma)$ entonces la función característica de X es

$$\phi_X(v) = \exp(i v^t \mu - \frac{1}{2} v^t \Sigma v), \quad (1.4)$$

$v \in \mathbb{R}^{m \times 1}$

Demostración.

Aquí

$$\phi_X(v) = E[\exp(i v^t X)] = \phi_{v^t X}(1),$$

El lado derecho denota la función característica de la variable aleatoria $v^t X$ evaluada en 1. Ya que X es $N_m(\mu, \Sigma)$ entonces $v^t X$ se distribuye como una normal univariada, $v^t X \sim N(v^t \mu, v^t \Sigma v)$ tal que

$$\phi_{v^t X}(1) = \exp(i v^t \mu - \frac{1}{2} v^t \Sigma v).$$

lo cual completa la demostración. □

Si u_i con $i = 1, 2, \dots, r$, son v.a.i.i.d con distribución normal estándar, si U es el vector de $r \times 1$ cuyos elementos son dichas variables.

$$\phi_U(v) = E[\exp(i v^t U)]$$

$$\begin{aligned}
&= \prod_{j=1}^r E\left[\exp(iv_j u_j)\right] && \text{(por} \\
&\text{independencia)} \\
&= \prod_{j=1}^r \exp\left(-\frac{1}{2}v_j^2\right) && \text{(por normalidad)} \\
&= \exp\left(-\frac{1}{2}\mathbf{v}^t\mathbf{v}\right)
\end{aligned}$$

Ahora si ponemos

$$\mathbf{X} = \mathbf{C}\mathbf{U} + \boldsymbol{\mu}$$

donde \mathbf{C} es una matriz de $m \times r$ de rango r tal que $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^t$, y $\boldsymbol{\mu} \in \mathbb{R}^{m \times 1}$. Entonces \mathbf{X} tiene función característica (1.4),

$$\begin{aligned}
E[\exp(i\mathbf{v}^t\mathbf{X})] &= E[\exp(i\mathbf{v}^t\mathbf{C}\mathbf{U})] \exp(i\mathbf{v}^t\boldsymbol{\mu}) \\
&= \phi_{\mathbf{U}}(\mathbf{C}^t\mathbf{v}) \exp(i\mathbf{v}^t\boldsymbol{\mu}) \\
&= \exp(-\frac{1}{2}\mathbf{v}^t\mathbf{C}\mathbf{C}^t\mathbf{v}) \exp(i\boldsymbol{\mu}^t\mathbf{v}) \\
&= \exp(i\mathbf{v}^t\boldsymbol{\mu} - \frac{1}{2}\mathbf{v}^t\boldsymbol{\Sigma}\mathbf{v}).
\end{aligned}$$

Vale comentar que podríamos haber definido la distribución normal multivariada $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ por medio de la transformación lineal sobre variables normal estándar independientes. Tal que dicha representación en la práctica es muy útil.

1.2.2 Formas lineales, Marginales y Condicionales.

Regresando a las propiedades de la distribución normal multivariada, los siguientes resultados muestran que cualquier transformación lineal de un vector normal tiene una distribución normal.

Teorema 3.8. Si \mathbf{X} es $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y \mathbf{B} es una matriz de dimensión $k \times m$, \mathbf{b} es un vector de $k \times 1$, entonces

$$\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b} \text{ es } N_k(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t).$$

Demostración.

El hecho de que \mathbf{Y} es normal k -variada es una consecuencia directa de la definición 3.5, puesto que todas las funciones lineales de los componentes de \mathbf{Y} son

funciones lineales de los componentes de X y estos son todos normales. La media y la covarianza de la matriz Y están claramente definidas.

□

Una propiedad importante de la distribución normal multivariada es que *todas* las distribuciones marginales son normales.

Teorema 3.9. Si X es $N_m(\mu, \Sigma)$, entonces la distribución marginal de cualquier subconjunto de $k < m$ componentes de X es normal k -variada.

Demostración.

Esto se sigue directamente de la definición, o del teorema 3.8. Por ejemplo, hacemos una partición de X , μ y Σ como

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

donde X_1 y μ_1 son de $k \times 1$ y Σ_{11} es de $k \times k$, poniendo

$$B = [I_k : 0] \quad \text{es de dimensión } k \times m \text{ y } b = \mathbf{0}$$

En el teorema 3.8 se muestra inmediatamente que X_1 es $N_m(\mu_1, \Sigma_{11})$.

□

Una consecuencia de este teorema es que la distribución marginal de cada componente de X es normal univariada. Lo inverso no es cierto en general; esto es, el hecho de que cada componente de un vector aleatorio no es (marginalmente) normal no implica que el vector tenga una distribución normal multivariada. [Esta es una de las razones por la cual el problema de *pruebas* (de hipótesis) de normalidad multivariada es hasta cierto punto un tanto complicado en la práctica].

Recordemos que la independencia de dos variables aleatorias implican que la covarianza entre ellas, si esta existe, es cero, pero que el caso contrario en general no es cierto. Esta característica, para la distribución normal multivariada, es como la muestra el siguiente resultado.

Teorema 3.10. Si X es $N_m(\mu, \Sigma)$ y X , μ y Σ es una partición como la siguiente:

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

donde X_1 y μ_1 son de $k \times 1$ y Σ_{11} es de $k \times k$, entonces los subvectores X_1 y X_2 son independientes si, y solo si $\Sigma_{12} = 0$.

Demostración.

Σ_{12} es la matriz de covarianzas entre los componentes de X_1 y los componentes de X_2 , de tal manera que la independencia de X_1 y X_2 implica que $\Sigma_{12} = 0$.

Sea $\Sigma_{12} = 0$, la función de densidad de X es,

$$\begin{aligned} f(\mathbf{X}) &= \frac{1}{(2\pi)^{p/2} \left| \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right|^{1/2}} e^{-\frac{1}{2} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix} \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix}} \\ &= \frac{1}{(2\pi)^{p/2} |\Sigma_{11}|^{1/2} |\Sigma_{22}|^{1/2}} e^{-\frac{1}{2} \left((X_1 - \mu_1)^t \Sigma_{11}^{-1} (X_1 - \mu_1) + (X_2 - \mu_2)^t \Sigma_{22}^{-1} (X_2 - \mu_2) \right)} \\ &= f(X_1) f(X_2) \end{aligned}$$

por lo tanto X_1, X_2 son independientes.

□

Este teorema se puede extender al caso donde se hacen particiones de \mathbf{X} en un número de subvectores. Lo importante de esto es que para determinar si dos subvectores de un vector distribuido normalmente son independientes es suficiente comprobar que la matriz de covarianzas entre los dos subvectores es cero.

Teorema 3.11. Si los vectores aleatorios X y Y de $m \times 1$ son independientes y $X + Y$ tienen una distribución normal m -variada, entonces X y Y se distribuyen como normal.

Demostración.

Para cada $\alpha \in \mathbb{R}^{m \times 1}$, la distribución de $\alpha^t(X + Y) = \alpha^t X + \alpha^t Y$ es normal (por el teorema 3.8, ya que $X + Y$ es normal). Puesto que $\alpha^t X$ y $\alpha^t Y$ son independientes, implica que ambas son normales, y por lo tanto X y Y son normal m -variada.

□

Una propiedad bien conocida de la distribución normal univariada es que las combinaciones lineales de variables normales independientes son normales. La generalización a la situación multivariada, es como se muestra a continuación.

Teorema 3.12. Si X_1, \dots, X_N son todas independientes, y X_i es $N_m(\mu_i, \Sigma_i)$ para $i = 1, \dots, N$, entonces para cualquier constante fija $\alpha_1, \dots, \alpha_N$,

$$\sum_{i=1}^N \alpha_i X_i \quad \text{es} \quad N_m \left(\sum_{i=1}^N \alpha_i \mu_i, \sum_{i=1}^N \alpha_i^2 \Sigma_i \right).$$

La demostración de este teorema se sigue de la definición 3.5, o por inspección de la función característica de $\sum_{i=1}^N \alpha_i X_i$.

Teorema 3.13. Si $X \sim N(\mu, \Sigma)$, $X \in R^{k \times 1}$, con $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, (X_1, X_2 son conjuntamente normales) $X_2 | X_1 \sim N(\mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (X_1 - \mu_1), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})$.

Demostración.

Sea $A = \begin{pmatrix} I & 0 \\ -\Sigma_{21} \Sigma_{11}^{-1} & I \end{pmatrix}$, es claro que $|A| = I$, si

$$W = A(X - \mu) = \begin{pmatrix} I & 0 \\ -\Sigma_{21} \Sigma_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix} = \begin{pmatrix} X_1 - \mu_1 \\ -\Sigma_{21} \Sigma_{11}^{-1} (X_1 - \mu_1) + (X_2 - \mu_2) \end{pmatrix}$$

Luego

$$W = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \end{pmatrix} \right)$$

$$f_{W_1 W_2}(W_1, W_2) =$$

$$\frac{1}{(2\pi)^{k_1/2} |\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}(W_1^t \Sigma_{11}^{-1} W_1)} \frac{1}{(2\pi)^{k_2/2} |\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}|^{1/2}} e^{-\frac{1}{2}(W_2^t (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})^{-1} W_2)}$$

$$f_{X_1, X_2}(X_1, X_2) = f_{W_1, W_2}(A(X - \mu)) |I|$$

$$= f_{W_1, W_2}(X_1 - \mu_1, (X_2 - \mu_2) - \Sigma_{21} \Sigma_{11}^{-1} (X_1 - \mu_1))$$

$$= \frac{1}{(2\pi)^{k_1/2} |\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}(W_1^t \Sigma_{11}^{-1} W_1)} \frac{1}{(2\pi)^{k_2/2} |\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}|^{1/2}} e^{-\frac{1}{2}(W_2^t (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})^{-1} W_2)}$$

$$f_{X_2 | X_1}(X_2 | X_1) =$$

$$\frac{1}{(2\pi)^{k_2/2} |\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{1/2}} e^{-\frac{1}{2}\left((X_2 - \mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)\right)' \left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1} \left((X_2 - \mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)\right)}$$

Como podemos ver $X_2 | X_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$.

□

Definición 3.6. A $Q = (X - \mu)' \Sigma^{-1}(X - \mu)$ se le denomina la forma cuadrática asociada a la función de densidad $f_X(X) = K e^{-\frac{1}{2}(X - \mu)' \Sigma^{-1}(X - \mu)}$

Teorema 3.14. Sea $X \sim N(\mu, \Sigma)$ con forma cuadrática Q , el vector de medias μ es aquel que da la solución al sistema de ecuaciones $\frac{\partial Q}{\partial X} = 0$.

Demostración.

La densidad la podemos poner como $f_X(X) = K e^{-\frac{1}{2}Q}$ el valor de X que maximiza $f_X(X)$ es el valor de X tal que $Q=0$, como $Q=(X - \mu)' \Sigma^{-1}(X - \mu)$ y es definida positiva, Q sólo puede ser cero en el punto $X - \mu = 0$, es decir el punto en donde $X = \mu$, por lo que podemos decir que μ es el punto que maximiza a $f_X(X)$.

La solución al sistema $\frac{\partial f_X(X)}{\partial X} = 0$ da el punto máximo de $f_X(X)$ y por lo mencionado anteriormente tal punto es μ . Pero como

$$\frac{\partial f_X(X)}{\partial X} = K e^{-\frac{1}{2}Q} \left(-\frac{1}{2} \right) \frac{\partial Q}{\partial X}$$

Se sigue que el vector que satisface $\frac{\partial f_X(X)}{\partial X} = 0$, es el mismo que satisface $\frac{\partial Q}{\partial X} = 0$. Por

tanto el vector μ es la solución al sistema $\frac{\partial Q}{\partial X} = 0$

□

Ejemplo 3.17. Sea x_1, x_2 , dos variables distribuidos conjuntamente como una normal multivariada con forma cuadrática

$$Q = x_1^2 + 2x_2^2 - x_1x_2 - 3x_1 - 2x_2 + 4$$

A partir de la forma cuadrática deseamos encontrar el vector de medias y la matriz de varianza covarianza. La forma cuadrática la podemos escribir como

$$Q = (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} - \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{X} - \mathbf{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$$

El único término que tiene términos de segundo grado es $\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X}$, así tenemos que

$$\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} = x_1^2 + 2x_2^2 - x_1x_2 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}' \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\text{así tenemos que } \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 2 \end{pmatrix} \text{ y } \boldsymbol{\Sigma} = \begin{pmatrix} \frac{8}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{4}{7} \end{pmatrix}$$

ahora sólo nos hace falta obtener la media, por el teorema anterior esta la podemos obtener resolviendo el sistema $\frac{\partial Q}{\partial \mathbf{X}} = 0$.

$$\text{Tenemos que } \frac{\partial Q}{\partial x_1} = 2x_1 - x_2 - 3 = 0$$

$$\frac{\partial Q}{\partial x_2} = -x_1 + 4x_2 - 2 = 0$$

La solución al sistema anterior es $x_2 = 1$, $x_1 = 2$ así tenemos que $\mu_1 = 2$ y $\mu_2 = 1$.