Probabilidad y estadística

MAT 041, Primer semestre

Francisco Cuevas Pacheco

11 de octubre de 2022

Contenidos

✓ Normal Multivariada

Definición (Normal Multivariada)

Diremos que el vector aleatorio $(X_1, \ldots, X_p)^\top$ *sigue una distribución normal multivariada con vector de media*

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$$

y matriz de covarianza

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}$$

si la función de densidad conjunta está dada por

$$f_{X_1,\ldots,X_p}(x_1, \ldots, x_p) = \frac{1}{(2\pi)^{p/2} det(\boldsymbol{\Sigma})^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

donde $\boldsymbol{x} = (x_1, \ldots, x_p)^\top$.

Observaciones

* Se utiliza la notación

$$(X_1, \ldots, X_p)^\top \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

* La componente k-ésima sigue una distribución normal (univariada):

$$X_k \sim \mathrm{N}(\mu_k, \sigma_{kk})$$

para cada $k = 1, \ldots, p$.

* La covarianza entre las componentes k-ésima y j-ésima es

$$\mathrm{cov}(X_k, X_j) = \sigma_{kj}$$

para cada $k, j = 1, \ldots, p$.

Caso Especial

Cuando $p = 2$, la función de densidad del vector bivariado

$$(X, Y)^\top \sim \mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

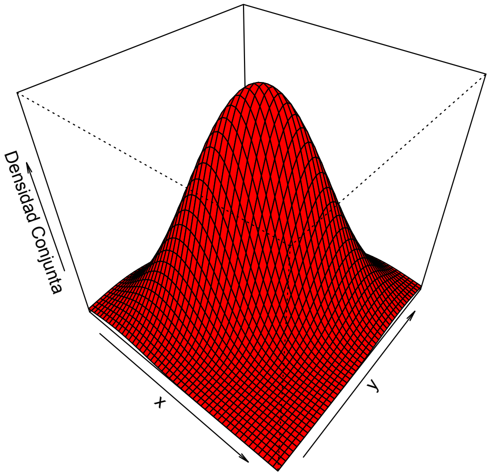
donde

$$\boldsymbol{\mu} = (\mu_X, \mu_Y)^\top \quad \text{y} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix},$$

está dada por

$$f_{XY}(x, y) = \frac{e^{-\frac{1}{2}\left(\frac{\sigma_Y^2(x-\mu_X)^2 + \sigma_X^2(y-\mu_Y)^2 - 2\sigma_{XY}(x-\mu_X)(y-\mu_Y)}{\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2}\right)}}{2\pi\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2}}$$

Distribución Normal Bivariada



Propiedad

Sea $(X, Y)^\top \sim \mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Considere la variable aleatoria

$$Z = aX + bY,$$

donde a y b son constantes. Entonces, $Z \sim \mathrm{N}(\mu_Z, \sigma_Z^2)$, donde

$$\mu_Z = E(Z) = E(aX + bY) = aE(X) + bE(Y) = a\mu_X + b\mu_Y$$

$$\sigma_Z^2 = \mathrm{var}(Z) = \mathrm{var}(aX + bY) = \mathrm{var}(aX) + \mathrm{var}(bY) + 2\,\mathrm{cov}(aX, bY)$$

$$= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\sigma_{XY}$$

Propiedad

Sea $(X, Y)^\top \sim \mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Entonces, X e Y son independientes si y solamente si

$$\sigma_{XY} = \mathrm{cov}(X, Y) = 0.$$

Ejercicio

Sean X e Y variables aleatorias independientes, cada una con distribución normal estándar. Definimos las variables aleatorias U y V:

$$U = 2X + Y, \qquad V = X - Y.$$

(a) Encuentre $E(U)$, $E(V)$ y $\mathrm{cov}(U, V)$.

(b) Calcule $P(U - V \geq 1)$.

Ejemplo

Se está estudiando la cantidad de agua que ingresa a una laguna la cual proviene de dos ríos principalmente, el río A y el río B, los caudales de ambos ríos se pueden modelar mediante una distribución normal bivariada, esto porque hay evidencia de que el caudal de ambos ríos está correlacionado, así:

$$(A, B) \sim \mathcal{N}\left(\begin{bmatrix} 5 \\ 8 \end{bmatrix}; \begin{pmatrix} 4 & 2 \\ 2 & 9 \end{pmatrix}\right)$$

¿Cual sería la distribución del caudal de agua que ingresa a la laguna?

Solución: Sean $A :=$ Caudal que ingresa por el río A y $B :=$ Caudal que entra por el río B.

$$A \sim \mathcal{N}(5, 2^2), \qquad B \sim \mathcal{N}(8, 3^2)$$

Sea $W = X + Y$ el caudal que ingresa a la laguna. Entonces

$$W \sim \mathcal{N}(\mu_W, \sigma_W^2)$$

con $\mu_W = 5 + 8 = 13$ y $\sigma_W^2 = 2^2 + 3^2 + 2 \cdot 2 = 17$

Distribución Condicional

Sea $(X, Y) \sim \mathcal{N}_2(\mu, \Sigma)$. Entonces la función de probabilidad condicional de X dado $Y = y$ viene dada por:

$$f_{X|Y=y}(x) = \frac{1}{\sqrt{2\pi\sigma_X^2(1-\rho^2)}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu_X - \rho(\sigma_X/\sigma_Y)(y-\mu_Y)}{\sigma_X\sqrt{1-\rho^2}}\right)^2\right],$$

es decir, la distribución condicional de un modelo normal sigue una distribución normal.

Distribución Condicional

Sea $(X, Y) \sim \mathcal{N}_2(\mu, \Sigma)$. Entonces la función de probabilidad condicional de X dado $Y = y$ viene dada por:

$$f_{X|Y=y}(x) = \frac{1}{\sqrt{2\pi\sigma_X^2(1-\rho^2)}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X-\rho(\sigma_X/\sigma_Y)(y-\mu_Y)}{\sigma_X\sqrt{1-\rho^2}}\right)^2\right],$$

es decir, la distribución condicional de un modelo normal sigue una distribución normal.

Para esta distribución la esperanza y la varianza vienen dadas por

$$E[X|Y=y] = \mu_X + \rho(\sigma_X/\sigma_Y)(y-\mu_Y)$$
$$V[X|Y=y] = \sigma_X^2(1-\rho^2)$$

Siguiendo el ejemplo anterior, calcule la distribución el río A dato que el caudal del río B es 7.

Siguiendo el ejemplo anterior, calcule la distribución el río A dato que el caudal del río B es 7.

Sabiendo que la distribución de A dado que B es igual a 7 es una distribución normal basta con identificar el valor de su media y su varianza.

$$E[A|B=7] = 5 + (2/(2*3))(2/3)(7-8) = 4.7$$
$$V[A|B=7] = 2^2(1-2^2/(4*9)) = 3.55$$

Así $A|_{B=7} \sim \mathcal{N}(4.7; 3.55)$

Teorema de transformación

Sea $X = (X_1, \ldots, X_p)$ un vector aleatorio p-dimensional con densidad f_X y considere el vector aleatorio $Y = (Y_1, \ldots, Y_p)$ dado por las biyecciones $Y_i = g_i(X_1, \ldots, X_p)$. Luego, la densidad conjunta del vector Y está dada por

$$f_Y(y_1, \ldots, y_p) = f_X(g_1^{-1}(y_1, \ldots, y_p), \ldots, g_p^{-1}(y_1, \ldots, y_p)) \left| J\left(\frac{x_1, \ldots, x_p}{y_1, \ldots, y_p}\right) \right|,$$

donde

$$J\left(\frac{x_1, \ldots, x_p}{y_1, \cdots, y_p}\right) = \begin{bmatrix} \frac{\partial x_1}{\partial y1} & \cdots & \frac{\partial x_1}{\partial y_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_p}{\partial y1} & \cdots & \frac{\partial x_p}{\partial y_p} \end{bmatrix}$$

es la matriz Jacobiana.

Ejemplo

Considere la siguiente función de densidad

$$f_{XY}(x,y) = \begin{cases} \frac{1}{2} & si -1 \leq x+y \leq 1, -1 \leq x-y \leq 1 \\ 0 & \textit{en otro caso}, \end{cases}$$

y considere la transformación $(U,V) = (X+Y, X-Y)$*. Encuentre* $f_{UV}(u,v)$

Solución: Notamos que $X = (U + V)/2$ e $Y = (U - V)/2$. En consecuencia, la matriz Jacobiana es

$$J\left(\frac{x_1, x_2}{y_1, y_2}\right) = \begin{bmatrix} \frac{\partial x_1}{\partial y1} & \frac{\partial x_1}{\partial y_p} \\ \frac{\partial x_2}{\partial y1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Luego $\left|J\left(\frac{x_1,x_2}{y_1,y_2}\right)\right| = 1/2$. En consecuencia se tiene que

$$f_{UV}(u,v) = f_{XY}((u+v)/2,(u-v)/2)(1/2) = \begin{cases} \frac{1}{4} & si -1 \leq u \leq 1, -1 \leq v \leq 1 \\ 0 & \text{en otro caso,} \end{cases}$$

Sea $X = (X_1, X_2)$ un vector aleatorio con densidad conjunta $f(x1, x2)$. Calcule la distribución de $Y_1 = X_1 + X_2$

Sea $X = (X_1, X_2)$ un vector aleatorio con densidad conjunta $f(x1, x2)$. Calcule la distribución de $Y_1 = X_1 + X_2$

Solución: Primero, consideramos una segunda coordenada $Y_2 = X_2$. Luego se tiene que $X_2 = Y_1 - Y_2$ y $X_1 = Y_2$

En consecuencia, la matriz Jacobiana es

$$J\left(\frac{x_1, x_2}{y_1, y_2}\right) = \begin{bmatrix} \frac{\partial x_{\mathbf{1}}}{\partial y1} & \frac{\partial x_{\mathbf{1}}}{\partial y_p} \\ \frac{\partial x_{\mathbf{2}}}{\partial y1} & \frac{\partial x_{\mathbf{2}}}{\partial y_{\mathbf{2}}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$