MAT-032: Distribuciones bivariadas

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Para un par de variables aleatorias X e Y se tiene la función de densidad conjunta como

$$f(x, y) = \mathsf{P}(X = x, Y = y)$$

Notación:

Escribiremos

$$\mathsf{P}(X = x, Y = y) = \mathsf{P}(\{X = x\} \cap \{Y = y\}).$$

Ejemplo:

Distribución bivariada con X e Y tomando valores 0 o 1

$X \setminus Y$	0	1	total
0	1/9	2/9	1/3
1	3/9	4/9	2/3
total	1/3	2/3	1

En este caso,

$$f(1, 1) = \mathsf{P}(X = 1, Y = 1) = \frac{4}{9}.$$

Definición 1 (Densidad conjunta):

Decimos que $f(x,y)$ es la función de densidad del vector aleatorio (X,Y) si:

(a) $f(x,y) \geq 0$ para todo (x,y).

(b) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$,

(c) Para cualquier conjunto $A \subset \mathbb{R}^2$,

$$\mathsf{P}((X,Y)\in A) = \int\int_A f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Observación:

Para el caso de distribuciones discretas, la definición es análoga.

Ejemplo:

Sea (X,Y) con densidad,

$$f(x,y) = \begin{cases} x+y, & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Entonces,

$$\begin{aligned} \int_0^1 \int_0^1 (x+y)\,\mathsf{d}x\,\mathsf{d}y &= \int_0^1 \Big[\int_0^1 x\,\mathsf{d}x\Big]\,\mathsf{d}y + \int_0^1 \Big[\int_0^1 y\,\mathsf{d}x\Big]\,\mathsf{d}y \\ &= \int_0^1 \frac{1}{2}\,\mathsf{d}y + \int_0^1 y\,\mathsf{d}y = \frac{1}{2} + \frac{1}{2} \\ &= 1 \end{aligned}$$

es decir, $f(x,y)$ es una densidad.

Definición 2 (Distribución acumulada conjunta):

La función de distribución acumulada conjunta (CDF) $F(x, y)$ es definida como:

$$F(x, y) = \mathsf{P}(X \leq x, Y \leq y).$$

Propiedades:

La CDF tiene las siguientes propiedades:

(a) $F(x, y)$ es función monótona creciente y continua a la derecha en (x, y).

(b) $0 \leq F(x, y) \leq 1$.

(c) $F(-\infty, y) = F(x, -\infty) = 0$.

(d) $F(+\infty, +\infty) = 1$.

Para el caso continuo, tenemos:

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(r,s)\,\mathsf{d}r\,\mathsf{d}s$$

Además, por el Teorema fundamental del Cálculo,

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

Resultado 1 (Densidad marginal):

Sea (X, Y) un vector aleatorio bivariado con densidad conjunta $f_{X,Y}(x, y)$. Entonces, las densidades marginales de X e Y, $f_X(x) = \mathsf{P}(X = x)$ y $f_Y(y) = \mathsf{P}(Y = y)$ son, respectivamente dadas por:

(a) Para el caso discreto:

$$f_X(x) = \sum_y f(x, y) = \sum_y \mathsf{P}(X = x, Y = y),$$
$$f_Y(y) = \sum_x f(x, y) = \sum_x \mathsf{P}(X = x, Y = y).$$

(b) Para variables continuas:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, \mathsf{d}y, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, \mathsf{d}x.$$

Ejemplo:

Suponga que

$$f_{X,Y}(x,y) = e^{-(x+y)}, \qquad x,y \geq 0.$$

De este modo,

$$f_X(x) = e^{-x} \int_0^\infty e^{-y} \, \mathsf{d}y = e^{-x}, \qquad x \geq 0.$$

Observación:

Para variables aleatorias continuas podemos usar la siguiente relación:

$$\begin{aligned} F_X(x) = \mathsf{P}(X \leq x) &= \int_{-\infty}^{x} f_X(t)\,\mathsf{d}t \\ &= \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f_{X,Y}(t,y)\,\mathsf{d}y\,\mathsf{d}t \\ &= F_{X,Y}(x,+\infty). \end{aligned}$$

Además,

$$f_X(x) = \frac{\mathsf{d}F_X(x)}{\mathsf{d}x},$$

y análogamente

$$F_Y(y) = F_{X,Y}(+\infty, y).$$

Definición 3 (Densidad condicional):

La función de densidad condicional para el caso discreto es definida como:

$$f_{X|Y}(x|y) = \mathsf{P}(X = x|Y = y) = \frac{\mathsf{P}(X = x, Y = y)}{\mathsf{P}(Y = y)}$$
$$= \frac{f_{X,Y}(x,y)}{f_Y(y)},$$

si $f_Y(y) > 0$. Para el caso continuo, tenemos

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \qquad f_Y(y) > 0.$$

Observación:

A pesar de la notación $f_{X|Y}(x|y)$ se debe destacar que $f_{X|Y}(x|y)$ es función de X.

Observación:

Evidentemente,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)},$$

y

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x} \\ &= \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)\,\mathrm{d}x}, \end{aligned}$$

que corresponde al Teorema de Bayes para densidades.

En efecto, la densidad condicional debe satisfacer las condiciones de una densidad.

$$\begin{aligned}\int_{-\infty}^{\infty} f_{X|Y}(x|y)\,\mathsf{d}x &= \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_Y(y)}\,\mathsf{d}x \\ &= \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathsf{d}x \\ &= \frac{f_Y(y)}{f_Y(y)} = 1.\end{aligned}$$

Además, es obvio que

$$f_{X|Y}(x|y) \geq 0.$$

Los momentos de una variable aleatoria condicional se definen de forma análoga:

$$\mathsf{E}(X|Y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,\mathsf{d}x,$$
$$\mathsf{E}(Y|X) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\,\mathsf{d}y,$$

y

$$\mathsf{var}(Y|X) = \mathsf{E}(Y^2|X) - \{\mathsf{E}(Y|X)\}^2.$$

Definición 4 (Independencia):

Sea (X, Y) un vector aleatorio con densidad conjunta $f(x, y)$ y densidades marginales $f_X(x)$, $f_Y(y)$, respectivamente. Entonces X e Y se dicen independientes si, para todo $x, y \in \mathbb{R}$,

$$f(x, y) = f_X(x) \cdot f_Y(y).$$

Observación:

Análogamente, si $F(x, y)$ representa la CDF del vector (X, Y) y $G(x)$, $H(y)$ las CDFs de X e Y, respectivamente. Entonces X e Y son independientes, si

$$F(x, y) = G(x) \cdot H(y).$$

Además, si X e Y son independientes, tenemos que:

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

Observación:

Diremos que $\boldsymbol{X} = (X_1, \ldots, X_k)^\top$ es un vector aleatorio k-dimensional.[1] Usaremos la notación $f(\boldsymbol{x})$ y $F(\boldsymbol{x})$ para la densidad y CDF del vector aleatorio $\boldsymbol{X}$. En este caso

$$F(\boldsymbol{x}) = \mathsf{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \mathsf{P}(X_1 \leq x_1, \ldots, X_k \leq x_k),$$

y asumiremos que existe f no negativa, tal que

$$F(\boldsymbol{x}) = \int_{-\infty}^{x} f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}, \qquad \boldsymbol{x} \in \mathbb{R}^k.$$

Además,

$$f(\boldsymbol{x}) = \frac{\partial^k F(\boldsymbol{x})}{\partial x_1 \cdots \partial x_k},$$

con $\int_{\mathbb{R}^k} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$.

[1] donde $X_1, \ldots, X_k$ son variables aleatorias.

Definición 5 (Variables IID):

Si $X_1, \dots, X_n$ son variables independientes y cada una tiene la misma distribución marginal con CDF F, decimos que $X_1, \dots, X_n$ son independientes e identicamente distribuídas (IID) y escribimos

$$X_1, \dots, X_n \sim F.$$

Si F tiene densidad F podemos escribir $X_1, \dots, X_n \sim f$.

Observación:

Cuando $X_1, \dots, X_n \overset{\text{IID}}{\sim} F$ también decimos que $X_1, \dots, X_n$ es una muestra aleatoria de tamaño n desde F (o f).[2]

Si $X_1, \dots, X_n$ son independientes, es fácil notar que

$$f(\boldsymbol{x}) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

[2] Y anotamos que $X_1, \dots, X_n$ es una m.a.(n) desde F.

A partir de ahora, consideraremos que $f(\cdot)$ es un miembro de una familia paramétrica:

$$\mathcal{P} = \{\mathsf{P}_\theta : \boldsymbol{\theta} \in \Theta\},$$

donde $\Theta \subset \mathbb{R}^p$ se denomina espacio paramétrico.

En cuyo caso escribimos $X \sim f(x;\theta)$. De este modo,

$$f(x_1,\dots,x_n;\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i;\boldsymbol{\theta}).$$

Para abreviar anotaremos

$$f(\boldsymbol{x};\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i;\boldsymbol{\theta}).$$

Evidentemente,

$$\log f(\boldsymbol{x};\boldsymbol{\theta}) = \sum_{i=1}^n \log f(x_i;\boldsymbol{\theta}).$$

EX UMBRA IN SOLEM

Ejemplo:

Sea $X_1, \dots, X_n$ m.a.(n) desde Exp(θ). De este modo, su distribución conjunta es dada por:

$$\begin{aligned} f(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \theta e^{-\theta x_i} \\ &= \theta^n \exp\Big(- \theta \sum_{i=1}^n x_i \Big) \\ &= \theta^n e^{-\theta n \overline{x}} \end{aligned}$$

Es decir, en este caso la densidad conjunta de $X_1, \dots, X_n$ es función de θ y $\overline{X}$.

Además,

$$\log f(\boldsymbol{x}; \theta) = n \log \theta - n\theta\overline{x}.$$

Recordatorio 1:

Sabemos que la función de densidad de $X \sim \mathsf{N}(\mu, \sigma^2)$, es dada por:

$$f(x;\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\Big\{ -\frac{1}{2}\Big(\frac{y-\mu}{\sigma}\Big)^2\Big\}$$

donde $\mu \in \mathbb{R}$ y $\sigma^2 > 0$. Además,

$$\mathsf{E}(X) = \mu, \qquad \mathsf{var}(X) = \sigma^2.$$

Evidentemente, si $Z \sim \mathsf{N}(0,1)$ entonces

$$\phi(z) = (2\pi)^{-1/2} \exp\big(-\tfrac{1}{2} z^2\big).$$

Propiedades:

(a) Si $X \sim \mathsf{N}(\mu,\sigma^2) \Leftrightarrow Z = \frac{X-\mu}{\sigma} \sim \mathsf{N}(0,1)$

(b) Si $X \sim \mathsf{N}(\mu,\sigma^2)$, entonces $Y = aX + b \sim \mathsf{N}(a\mu + b, a^2\sigma^2)$.

Sea $\boldsymbol{X} = (X_1, \dots, X_p)^\top$, la distribución normal multivariada es caracterizada por un vector de medias $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^\top$ y una matriz de covarianzas $\boldsymbol{\Sigma} = (\sigma_{ij}) \geq \boldsymbol{0}$. En cuyo caso escribimos $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Considere $Z_1, \dots, Z_p$ variables IID desde $\mathsf{N}(0, 1)$ y suponga $\boldsymbol{Z} = (Z_1, \dots, Z_p)^\top$. Entonces,

$$
\begin{aligned}
f_Z(\boldsymbol{z}) &= \prod_{i=1}^{p} f_{Z_i}(z_i) = \prod_{i=1}^{p} (2\pi)^{-1/2} \exp\big(-\tfrac{1}{2} z_i^2\big) \\
&= (2\pi)^{-p/2} \exp\Big(-\frac{1}{2} \sum_{i=1}^{p} z_i^2\Big) \\
&= (2\pi)^{-p/2} \exp\big(-\tfrac{1}{2} \|\boldsymbol{z}\|^2\big),
\end{aligned}
$$

y anotamos $\boldsymbol{Z} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I})$.

Mas generalmente, $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ tiene densidad

$$f_X(\boldsymbol{x}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\big\{ -\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big\}.$$

Dado que $\boldsymbol{\Sigma}$ es simétrica y definido positiva tenemos que $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^\top$ esto lleva al resultado siguiente.

Resultado 2:

Si $\boldsymbol{Z} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I})$, entonces

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{Z} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Asimismo si $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces

$$\boldsymbol{B}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I}).$$

Ejemplo:

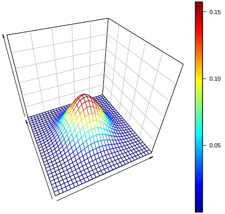
Sea $\boldsymbol{Z} \sim \mathsf{N}_2(\mathbf{0}, \boldsymbol{\Sigma})$, donde

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad -1 < \rho < 1.$$

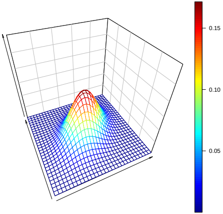
En cuyo caso, la función de densidad adopta la forma:

$$f(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big\{ -\frac{1}{2(1-\rho^2)}(x_1^2 + x_2^2 - 2\rho x_1 x_2)\Big\}.$$

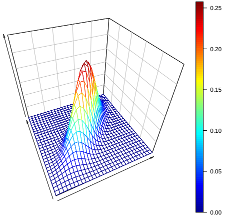
En la siguiente slide se presenta la densidad para $\rho = 0.0, 0.4$ y 0.8.



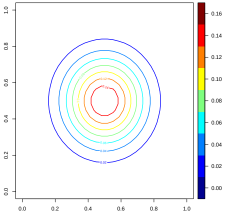
(a) $\rho = 0.0$



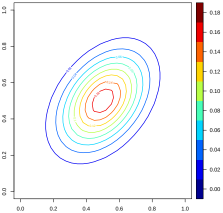
(b) $\rho = 0.4$



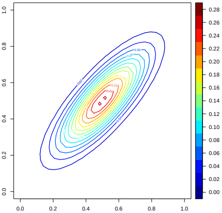
(c) $\rho = 0.8$



(d) contornos



(e) contornos



(f) contornos

Resultado 3:

Suponga $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b} \sim \mathsf{N}_k(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top),$$

donde $\boldsymbol{A} \in \mathbb{R}^{k\times p}$ con $\operatorname{rg}(\boldsymbol{A}) = k$ y $\boldsymbol{b} \in \mathbb{R}^k$.

Resultado 4:

Suponga $\boldsymbol{Z} = (Z_1, \dots, Z_p)^\top \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I})$,[3] entonces

$$U = \boldsymbol{Z}^\top \boldsymbol{Z} = \sum_{i=1}^p Z_i^2 \sim \chi^2(p),$$

donde $\chi^2(p)$ denota una distribución chi-cuadrado con p grados de libertad.[4]

[3] Es decir, $Z_1, \dots, Z_p$ son variables aleatorioas IID

[4] $U \sim \chi^2(p) \stackrel{\mathrm{d}}{=} \mathsf{Gama}(p/2, 1/2)$.

Resultado 5:

Sea $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y considere la partición:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Entonces,

(a) La distribución marginal de $\boldsymbol{X}_i \sim \mathsf{N}_{p_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ii})$, $i = 1, 2$.

(b) La distribución condicional de $\boldsymbol{X}_2 | \boldsymbol{X}_1 = \boldsymbol{x}_1$ es

$$\boldsymbol{X}_2 | \boldsymbol{X}_1 = \boldsymbol{x}_1 \sim \mathsf{N}_{p_2}\big(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{21}\big).$$

(c) $U = (\boldsymbol{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2(p)$.