

# Distribuciones conjuntas de probabilidad

## 1.1. Distribución de probabilidad bivariadas

**Definición 1.1.** Sean  $X$  e  $Y$  dos variables aleatorias discretas. La probabilidad de que  $X = x$  e  $Y = y$  está determinada por la función de probabilidad bivariada

$$p(x, y) = P(X = x, Y = y),$$

en donde  $P(x, y) \geq 0$  para toda  $x, y$ , de  $X, Y$ , y  $\sum_x \sum_y p(x, y) = 1$ .

La **función de distribución acumulativa bivariada** es la probabilidad conjunta de que  $X \leq x$  y  $Y \leq y$ , dada por

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_i \leq y} p(x_i, y_i).$$

La **función de distribución trinomial** viene dado por:

$$p(x, y, n, p_1, p_2) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}$$

y su generalización llamada **función de distribución multinomial** viene dada por:

$$p(x_1, x_2, \dots, x_{k-1}; n, p_1, p_2, \dots, p_{k-1}) = \frac{n!}{x_1!x_2!\dots x_{k-1}!} p_1^{x_1} p_2^{x_2} \dots p_{k-1}^{x_{k-1}}, \quad x_i = 0, 1, \dots, n \text{ para } i = 1, 2, \dots, k$$

en donde  $x_k = n - x_1 - x_2 - \dots - x_{k-1}$  y  $p_k = 1 - p_1 - p_2 - \dots - p_{k-1}$ .

**Definición 1.2.** Sean  $X$  e  $Y$  dos variables aleatorias continuas. Si existe una función  $f(x, y)$  tal que la probabilidad conjunta:

$$P(a < X < b, c < Y < d) = \int_a^b \int_c^d f(x, y) dy dx$$

para cualquier valor de  $a, b, c$  y  $d$  en donde  $f(x, y) \geq 0$ ,  $-\infty < x, y < \infty$  y  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1$ , entonces  $f(x, y)$  es la función de densidad de probabilidad bivariada de  $X$  e  $Y$ .

La **función de distribución bivariada acumulativa** de  $X$  e  $Y$  es la probabilidad conjunta de que  $X \leq x$  e  $Y \leq y$ , dada por:

$$P(X \leq x, Y \leq y) = F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dv du.$$

Por lo tanto, la función de densidad bivariada se encuentra diferenciando  $F(x, y)$  con respecto a  $x$  e  $y$ ; es decir,

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

## 1.2. Distribuciones marginales de probabilidad

Es posible determinar varias distribuciones marginales para cualquier distribución de probabilidad que contenga más de dos variables aleatorias.

**Definición 1.3.** Sean  $X$  e  $Y$  dos variables aleatorias discretas con una función de probabilidad conjunta  $p(x, y)$ . Las funciones marginales de probabilidad de  $X$  y de  $Y$  están dadas por

$$p_X(x) = \sum_y p(x, y) \quad \text{y} \quad p_Y(y) = \sum_x p(x, y),$$

respectivamente.

**Definición 1.4.** Sean  $X$  e  $Y$  dos variables aleatorias continuas con una función de densidad de probabilidad conjunta  $f(x, y)$ . Las funciones de densidad de probabilidad de  $X$  e  $Y$  están dadas por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy \quad \text{y} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx,$$

respectivamente.

Para variables aleatorias continuas conjuntas, si se conoce **la función de distribución acumulativa**  $F(x, y)$ , las distribuciones acumulativas marginales de  $X$  e  $Y$  se obtienen de la siguiente forma:

$$P(X \leq x) = F_X(x) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(t, y) \, dy dt, \quad \text{y} \quad F_X(x) = \int_{-\infty}^x f_X(t) \, dt = F(x, \infty)$$

De manera similar

$$P(Y \leq y) = F_Y(y) = \int_{-\infty}^y \int_{-\infty}^{\infty} f(x, t) \, dx dt = \int_{-\infty}^y f_Y(t) \, dt = F(\infty, y).$$

Así puede determinarse la distribución acumulativa marginal de  $X$  dejando que  $Y$  tome un valor igual al límite superior de la función de distribución conjunta de  $X$  e  $Y$ .

### 1.3. Valores esperados y momentos para distribuciones bivariadas

**Definición 1.5.** Sean  $X$  e  $Y$  dos variables aleatorias que se distribuyen conjuntamente. El valor esperado de una función de  $X$  y de  $Y$ ,  $g(x, y)$ , se define como

$$E[g(X, Y)] = \sum_x \sum_y g(x, y) p(x, y)$$

si  $X$  e  $Y$  son discretas, o

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

si  $X$  e  $Y$  son continuas, en donde  $p(x, y)$  y  $f(x, y)$  son las funciones de probabilidad y de densidad de probabilidad conjuntas, respectivamente.

Como consecuencia de la definición anterior, el **r-ésimo momento de  $X$  alrededor del cero** es

$$E(X^r) = \int_{-\infty}^{\infty} x^r f(x, y) dy dx = \int_{-\infty}^{\infty} x^r f_X(x) dx.$$

De manera similar

$$E(Y^r) = \int_{-\infty}^{\infty} y^r f_Y(y) dy.$$

El **r y s-ésimo momento producto de  $X$  e  $Y$  alrededor del origen** es:

$$E(X^r Y^s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s f(x, y) dy dx.$$

y **alrededor de las medias** es

$$E[(X - \mu_X)^r (Y - \mu_Y)^s] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^r (y - \mu_Y)^s f(x, y) dy dx.$$

De particular importancia es el momento producto alrededor de las medias cuando  $r = s = 1$ . Este momento producto recibe el nombre de **covarianza de  $X$  e  $Y$** , y se encuentra definido por

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

Al igual que la varianza, que es una medida de dispersión de una variable aleatoria, la covarianza es una medida de la variabilidad conjunta de  $X$  e  $Y$ . De esta forma, la covarianza es una medida de asociación entre los valores de  $X$  e  $Y$  y sus respectivas dispersiones.

Desarrollando el miembro derecho de  $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$  se tiene

$$E[(X - \mu_X)(Y - \mu_Y)] = E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] = E(XY) - \mu_X\mu_Y;$$

de esta forma

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Si la covarianza de  $X$  e  $Y$  se divide por el producto de las desviaciones estándar de  $X$  e  $Y$ , el resultado es una cantidad sin dimensiones que recibe el nombre de **coeficiente de correlación** y que se denota por  $\rho(X, Y)$ :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Se puede demostrar que el coeficiente de correlación se encuentra contenido en el intervalo  $-1 \leq \rho \leq 1$ . De hecho  $\rho$  es la covarianza de dos variables aleatorias estandarizadas  $X'$  e  $Y'$  en donde

$$X' = \frac{X - \mu_X}{\sigma_X} \quad \text{e} \quad Y' = \frac{Y - \mu_Y}{\sigma_Y}.$$

Esto significa que el coeficiente de correlación es sólo una medida estandarizada de la asociación lineal que existe entre las variables aleatorias  $X$  e  $Y$  en relación con sus dispersiones. El valor  $\rho = 0$  indica la ausencia de cualquier asociación lineal, mientras que los valores  $-1$  y  $1$  indican relaciones lineales perfectas negativas y positivas, respectivamente.

#### 1.4. Variables aleatorias estadísticamente independientes

**Definición 1.6.** Sean  $X$  e  $Y$  dos variables aleatorias con una distribución conjunta. Se dice que  $X$  e  $Y$  son estadísticas independientes si y sólo si,

$$p(x, y) = p_X(x)p_Y(y) \text{ si } X \text{ e } Y \text{ son discretas}$$

o bien

$$f(x, y) = f_X(x)f_Y(y) \text{ si } X \text{ e } Y \text{ son continuas,}$$

para todo  $x$  e  $y$ , en donde  $p(x, y)$  y  $f(x, y)$  son las funciones bivariadas de probabilidad y de densidad de probabilidad, respectivamente, y en donde  $p_X(x)$ ,  $p_Y(y)$ ,  $f_X(x)$  y  $f_Y(y)$  son las funciones de probabilidad marginal o de densidad de probabilidad marginal apropiadas.

Se desprende de esta definición que si  $X$  e  $Y$  son estadísticamente independientes, la probabilidad conjunta

$$\begin{aligned} P(a < X < b, c < Y < d) &= \int_a^b \int_c^d f(x, y) \, dy \, dx \\ &= \int_a^b \int_c^d f_X(x)f_Y(y) \, dy \, dx \\ &= \int_a^b f_X(x) \, dx \int_c^d f_Y(y) \, dy \\ &= P(a < X < b)P(c < Y < d). \end{aligned}$$

Por la misma condición,

$$\begin{aligned}
 E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y) dy dx \\
 &= \int_a^b \int_c^d xyf_X(x)f_Y(y) dy dx \\
 &= \int_c^d xf_X(x) dy \int_a^b yf_Y(y) dx \\
 &= E(X)E(Y).
 \end{aligned}$$

Si  $X$  e  $Y$  son estadísticamente independientes, entonces  $\text{Cov}(X, Y) = \rho(X, Y) = 0$ . Sin embargo la proposición inversa no es cierta.

Sean  $X$  e  $Y$  dos variables aleatorias continuas con una función de densidad conjunta de probabilidad  $f(x, y)$ . El **valor esperado de una función lineal** de  $X$  e  $Y$  es

$$\begin{aligned}
 E(aX + bY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by)f(x, y) dy dx \\
 &= a \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y) dy dx + b \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x, y) dy dx \\
 &= aE(X) + bE(Y).
 \end{aligned}$$

para cualquier valor de las constantes  $a$  y  $b$ .

La varianza de una función lineal de  $X$  e  $Y$  es

$$\begin{aligned}
 \text{Var}(aX + bY) &= E(aX + bY)^2 - E^2(aX + bY) \\
 &= E(a^2X^2 + 2abXY + b^2Y^2) - [aE(X) + bE(Y)]^2 \\
 &= a^2E(X^2) + 2abE(XY) + b^2E(Y^2) - a^2E^2(X) - 2abE(X)E(Y) - b^2E^2(Y) \\
 &= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).
 \end{aligned}$$

Además si  $X$  e  $Y$  son estadísticamente independientes,

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

La generalización de estos resultados a  $n$  variables aleatorias se hace por inducción y se establece en el siguiente teorema:

**Teorema 1.1.** Sean  $X_1, X_2, \dots, X_n$   $n$  variables aleatorias con una función de densidad conjunta de probabilidad  $f(x_1, x_2, \dots, x_n)$ . Entonces

$$\begin{aligned}
 E \left[ \sum_{i=1}^n a_i X_i \right] &= \sum_{i=1}^n [a_i E(X_i)] \\
 \text{Var} \left[ \sum_{i=1}^n a_i X_i \right] &= \sum_{i=1}^n a_i^2 \text{Var}(X_i) + \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j)
 \end{aligned}$$

para cualquier constante  $a_i$ ,  $i = 1, 2, \dots, n$ .

### 1.5. Distribuciones de probabilidad condicional

Considere la función

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

en donde  $f_Y(y)$  es la densidad marginal de  $Y$ . Si se mantiene constante a la variable aleatoria  $Y$  en el valor observado y de manera tal que  $f_Y(y) > 0$ , entonces  $\frac{f(x, y)}{f_Y(y)}$  define una función no negativa de  $X$  cuya integral es 1, dado que por definición

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

De esta forma,  $\frac{f(x, y)}{f_Y(y)}$  es una función de densidad de probabilidad y la probabilidad de que  $a < X < b$ , dado que el nivel de concentración de  $Y$  es  $y$ , está dada por:

$$P(a < X < b/y) = \int_a^b \frac{f(x, y)}{f_Y(y)} dx.$$

**Definición 1.7.** Sean  $X$  e  $Y$  dos variables aleatorias con una función de densidad conjunta de probabilidad  $f(x, y)$ . La función de densidad de probabilidad condicional de la variable aleatoria  $X$ , denotada por  $f(x|y)$ , para un valor fijo  $y$  de  $Y$ , está definida por

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

En donde  $f_Y(y)$  es la función de densidad de probabilidad de  $Y$  de manera tal que  $f_Y(y) > 0$ .

De manera análoga, la función de densidad de probabilidad condicional de  $Y$  para un valor fijo  $x$  de  $X$  se define como

$$f(y|x) = \frac{f(x, y)}{f_X(x)} \quad f_X(x) > 0.$$

En donde  $f_X(x)$  es la densidad marginal de  $X$ .

Nótese que si la densidad condicional  $f(x|y)$  por ejemplo, no contiene a  $y$ , entonces  $X$  es estadísticamente independientes de  $Y$ . Esto es, si  $X$  son estadísticamente independientes entonces

$$f(x, y) = f_X(x)f_Y(y)$$

y

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

De manera similar, si

$$f(x, y) = f_X(x)f_Y(y),$$

entonces

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

Los **valores esperados condicionales** se definen de manera análoga a la señalada en la definición 6.5. Por ejemplo, los valores esperados condicionales de  $X$  puesto que  $Y = y$ ,  $y$  de  $Y$ , ya que  $X = x$ , se definen como

$$E(X|y) = \int_{-\infty}^{\infty} xf(x|y) dx.$$

Y

$$E(Y|x) = \int_{-\infty}^{\infty} y f(y|x) dy.$$

De manera similar,

$$Var(X|y) = E(X^2|y) - E^2(X|y).$$

Y

$$Var(Y|x) = E(Y^2|x) - E^2(Y|x).$$

En donde

$$E(X^2|y) = \int_{-\infty}^{\infty} x^2 f(x|y) dx \quad y \quad E(Y^2|x) = \int_{-\infty}^{\infty} y^2 f(y|x) dy.$$

### 1.6. Análisis bayesiano: las distribuciones a priori y a posteriori.

**Teorema 1.2.** Sea  $p_Y(y)$  o  $f_Y(y)$  la función de probabilidad o de densidad de probabilidad a priori de  $Y$ , respectivamente, y sea  $f(x|y)$  la función de verosimilitud. Entonces la probabilidad a posteriori o función de probabilidad a posteriori de  $Y$  dada la evidencia muestral  $x$ , es

$$p(y|x) = \frac{f(x|y)p_Y(y)}{\sum_Y f(x|y)p_Y(y)} \text{ si } Y \text{ es discreta ,}$$

$$p(y|x) = \frac{f(x|y)p_Y(y)}{\int_{-\infty}^{\infty} f(x|y)p_Y(y) dy} \text{ si } Y \text{ es continua .}$$

Es interesante notar que el denominador de es la función de densidad de probabilidad marginal o no condicional de  $X$ ; esto es,

$$f_X(x) = \sum_Y f(x|y)p_Y(y)$$

o

$$f_X(x) = \int_Y f(x|y)f_Y(y) dy.$$

Dependiendo de cuando  $Y$  es discreta o continua. Además, el numerador es el producto de la función de verosimilitud y la función de probabilidad a priori  $y$ , de ésta manera, es la probabilidad conjunta de  $X$  e  $Y$  expresado como

$$f(x, y) = f(x|y)p_Y(y) \text{ si } Y \text{ es discreta,}$$

o

$$f(x, y) = f(x|y)f_Y(y) dy \text{ si } Y \text{ es continua.}$$

Nótese que para la penúltima ecuación la función  $f(x, y)$  es una mezcla bivariada de una variable aleatoria continua y otra discreta.

### 1.7. La distribución normal bivariada

**Definición 1.8.** Se dice que las variables aleatorias  $X$  e  $Y$  tienen una distribución normal bivariada si su función de densidad conjunta de probabilidad está dada por

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

con  $-\infty < x, y < \infty$ . Donde

$$\mu_X = E(X), \quad \mu_Y = E(Y), \quad \sigma_X^2 = \text{Var}(X), \quad \sigma_Y^2 = \text{Var}(Y),$$

y  $\rho$  es el coeficiente de correlación de  $X$  e  $Y$ .

Es interesante notar que, a pesar de que  $\rho = 0$  es una condición necesaria de independencia, para la distribución normal bivariada también es una condición suficiente. Eso es, si  $\rho = 0$  entonces

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2} \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - \frac{1}{2} \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}} \\ &= f_X(x)f_Y(y). \end{aligned}$$

Donde,  $f_X(x)$  e  $f_Y(y)$  son las densidad normales univariadas de  $X$  e  $Y$ , respectivamente.