Introduction to Statistics (MAT 283)

Dipti Dubey

Department of Mathematics
Shiv Nadar University

Table of Contents

CONDITIONAL DISTRIBUTION :

Let X and Y be any two random variables with joint pdf (or pmf) f and marginals f_X and f_Y. The conditional probability density function (or pmf) g of X, given (the event) $Y = y$, is defined as

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

provided $f_Y(y) > 0$.

CONDITIONAL DISTRIBUTION :

Let X and Y be any two random variables with joint pdf (or pmf) f and marginals f_X and f_Y. The conditional probability density function (or pmf) g of X, given (the event) $Y = y$, is defined as

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

provided $f_Y(y) > 0$.

Similarly, the conditional probability density function (or pmf) h of Y, given (the event) $X = x$, is defined as

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

provided $f_X(x) > 0$

Example: Let X and Y be discrete random variables with joint probability mass function

$$f(x,y) = \begin{cases} \frac{1}{21}(x+y) & \text{if} \quad x = 1,2,3, \; y = 1,2 \\ \\ 0 & \text{otherwise,} \end{cases}$$

What is the conditional probability mass function of X, given $Y = 2$?

INDEPENDENCE OF RANDOM VARIABLES:

Let X and Y be any two random variables with joint cdf F and marginals F_X and F_Y. The random variables X and Y are independent if and only if

$$F(x, y) = F_X(x)F_Y(y),$$

for all $(x, y) \in \mathbb{R}^2$.

Theorem:
(a) A necessary and sufficient condition for random variables X and Y of the discrete type to be independent is that

$$P(X = x_i, Y = y_i) = P(X = x_i)P(Y = y_i)$$

for all $(x_i, y_) \in R_X \times R_Y$.

(b) Two random variables X and Y of the continuous type are independent if and only if

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

for all $(x, y) \in \mathbb{R}^2$, where f , f_X, f_Y, respectively, are the joint and marginal pdfs of X and Y.

Example: Let X and Y be continuous random variables with joint pdf

$$f(x, y) = \begin{cases} e^{-(x+y)} & \text{if} \quad 0 < x, y < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

Are X and Y independent?

Theorem: Let X and Y be independent random variables and $\phi, \psi : \mathbb{R} \to \mathbb{R}$ are Borel measurable functions. Then the random variables $\phi(X)$ and $\psi(Y)$ are also independent.

Proof: We have

$$
\begin{aligned}
P(\phi(X) \leq x, \psi(Y) \leq y) &= P(X \in \phi^{-1}(-\infty, x], Y \in \psi^{-1}(-\infty, y]) \\
&= P(X \in \phi^{-1}(-\infty, x]) \; P(Y \in \psi^{-1}(-\infty, y]) \\
&= P(\phi(X) \leq x) \; P(\psi(Y) \leq y).
\end{aligned}
$$

Hence the proof.

IID Random Variables :

The random variables X and Y are said to be independent and identically distributed (IID) if and only if they are independent and have the same distribution.

Table of Contents

PRODUCT MOMENT ABOUT THE ORIGIN :

Let X and Y be any two random variables with joint pdf or pmf f. The product moment of X and Y about the origin, denoted by $E(XY)$, is defined as

$$E(XY) = \begin{cases} \sum_{x \in R_X} \sum_{y \in R_Y} xy\ f(x, y) & \text{if } X,\ Y \text{ are discrete} \\ \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\ f(x, y) dxdy & \text{if } X,\ Y \text{ are continuous,} \end{cases}$$

provided $E(XY) < \infty$.

COVARIANCE: The covariance between X and Y , denoted by $Cov(X, Y)$ (or σ_{XY}), is defined as

$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

where μ_X and μ_Y are mean of X and Y respectively.

COVARIANCE: The covariance between X and Y , denoted by $Cov(X, Y)$ (or σ_{XY}), is defined as

$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

where μ_X and μ_Y are mean of X and Y respectively.

• For an arbitrary random variable, the product moment and covariance may or may not exist. Further, note that unlike variance, the covariance between two random variables may be negative.

Theorem: Let X and Y be any two random variables. Then

$$Cov(X, Y) = E(XY) - E(X)E(Y).$$

Proof:

$$\begin{aligned}
Cov(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) \\
&= E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\
&= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\
&= E(XY) - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y \\
&= E(XY) - \mu_X \mu_Y \\
&= E(XY) - E(X)E(Y).
\end{aligned}$$

Corollary: $Cov(X, X) = \sigma_X^2$.
Proof:

EXAMPLE: Let X and Y be discrete random variables with joint pmf

$$f(x,y) = \begin{cases} \frac{x+2y}{18} & \text{for} \quad x = 1,2; y = 1,2 \\ \\ 0 & \qquad \text{otherwise,} \end{cases}$$

What is the covariance σ_{XY} between X and Y.

Theorem: If X and Y are any two random variables and a, b, c, and d are real constants, then

$$Cov(aX + b, cY + d) = ac\ Cov(X, Y).$$

Theorem: If X and Y are any two random variables and a, b, c, and d are real constants, then

$$Cov(aX + b, cY + d) = ac\ Cov(X, Y).$$

Proof: We have

$$\begin{aligned} Cov(aX + b, cY + d) &= E((aX + b)(cY + d)) - E(aX + b)E(cY + d) \\ &= E(acXY + adX + bcY + bd) \\ &\quad - (aE(X) + b)(cE(Y) + d) \\ &= acE(XY) + adE(X) + bcE(Y) + bd \\ &\quad - [acE(X)E(Y) + adE(X) + bcE(Y) + bd] \\ &= ac[E(XY) - E(X)E(Y)] \\ &= ac\ Cov(X, Y). \end{aligned}$$

EXAMPLE: If the product moment of X and Y is 3 and the mean of X and Y are both equal to 2, then what is the covariance of the random variables $2X + 10$ and $-\frac{5}{2}Y + 3$?

Theorem: If X and Y are independent random variables, then

$$E(XY) = E(X)E(Y).$$

Proof: Let us assume that X and Y are continuous. Therefore

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \ f(x,y)dxdy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \ f_X(x)f_Y(y)dxdy \\
&= \left(\int_{-\infty}^{\infty} x \ f_X(x)dx \right) \ \left(\int_{-\infty}^{\infty} y \ f_Y(y)dy \right) \\
&= E(X)E(Y).
\end{aligned}$$

If X and Y are discrete, then replace the integrals by appropriate sums to prove the same result.

Corollary: If X and Y are independent random variables, then $Cov(X, Y) = 0$.

EXAMPLE: Let the random variables X and Y have the joint pmf

$$f(x,y) = \begin{cases} \frac{1}{4} & \text{if} \quad (x,y) = \{(0,1),\ (0,-1),\ (1,0),\ (-1,0)\} \\ \\ 0 & \quad \text{otherwise,} \end{cases}$$

What is the covariance of X and Y ? Are the random variables X and Y independent?

	Y=-1	Y=0	Y=1	P(X=x)
X= -1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
X= 0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{2}{4}$
X= 1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
P(Y=y)	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	

Theorem: Let X and Y be any two random variables and let a and b be any two real numbers. Then

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y).$$

Theorem: Let X and Y be any two random variables and let a and b be any two real numbers. Then

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y).$$

Proof

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

- In case of three random variables $X,\ Y,\ Z$, we have

$$\begin{aligned} Var(X+Y+Z) &= Var(X) + Var(Y) + Var(Z) \\ &\quad + 2Cov(X,Y) + 2Cov(Y,Z) + 2Cov(Z,X) \end{aligned}$$