Joint distribution of a vector of random variables

- Suppose that X_1, ..., X_k are k random variables defined on the same sample space. Then the (joint) distribution of $(X_1, \ldots, X_k)$ is the probability function Q defined by

$$Q(A) = P((X_1, \ldots, X_k) \in A)$$

for $A \in \mathcal{B}(R^k)$, where $\mathcal{B}(R^k)$ is the σ-field generated by open sets in R^k.

- Joint cumulative distribution function (joint CDF). Suppose that $(X_1, \ldots, X_k)$ is a vector of k random variables defined on the same sample space. The joint distribution of $(X_1, \ldots, X_k)$ can be characterized by $F_{X_1,\ldots,X_k}$: the joint CDF of $(X_1, \ldots, X_k)$, which is defined as follows:

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = P(\{X_1 \leq x_1\} \cap \cdots \cap \{X_k \leq x_k\})$$

for $(x_1, \ldots, x_k) \in R^k$.

  - Suppose that X and Y are two random variables defined on the same sample space. The joint CDF of (X, Y), denoted by $F_{X,Y}$, is a function on R^2 defined by

$$F_{X,Y}(x, y) = P(X \leq x \text{ and } Y \leq y)$$

  for $(x, y) \in R^2$.

- Properties of a bivariate CDF. Suppose that X and Y are two random variables defined on the same sample space. Let $F_{X,Y}$, F_X, and F_Y be the CDFs of (X, Y), X, and Y respectively. For a fixed x, let $H(y) = F_{X,Y}(x, y)$ for $y \in (-\infty, \infty)$, then

  (a) H is increasing.

  (b)

$$\lim_{y \to -\infty} H(y) = \lim_{y \to -\infty} F_{X,Y}(x, y) = 0$$
$$\lim_{y \to \infty} H(y) = \lim_{y \to \infty} F_{X,Y}(x, y) = F_X(x).$$

  (c) H is right continuous.

  Similarly, for a fixed y, let $G(x) = F_{X,Y}(x, y)$ for $x \in (-\infty, \infty)$, then

  (d) G is increasing.

  (e)

$$\lim_{x \to -\infty} G(x) = \lim_{x \to -\infty} F_{X,Y}(x, y) = 0$$
$$\lim_{x \to \infty} G(x) = \lim_{x \to \infty} F_{X,Y}(x, y) = F_Y(y).$$

  (f) G is right continuous.

- When the joint CDF of (X, Y) is known, the distribution of X and the distribution of Y can be obtained. The CDF of X (or Y) is also called the marginal CDF of X (or Y).

- Example 1. Suppose that (X, Y) has joint CDF $F_{X,Y}$ given by

$$F_{X,Y}(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}) & \text{if } x \geq 0 \text{ and } y \geq 0; \\ 0 & \text{otherwise.} \end{cases}$$

  Find the (marginal) CDF of X and $P(0 < X \leq 1)$.

  Sol. Let F_X be the CDF of X, then

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y) = \begin{cases} \lim_{y\to\infty}(1 - e^{-x})(1 - e^{-y}) = 1 - e^{-x} & \text{if } x \geq 0; \\ \lim_{y\to\infty} 0 = 0 & \text{if } x < 0. \end{cases}$$

$$P(0 < X \leq 1) = F_X(1) - F_X(0) = 1 - e^{-1} - (1 - e^0) = 1 - e^{-1}.$$

- Suppose that (X, Y) has joint CDF F. Then for real numbers a, b, c, d such that $a < b$ and $c < d$,

$$P(a < X \leq b \text{ and } c < Y \leq d) = F(b, d) - F(b, c) - F(a, d) + F(a, c). \quad (1)$$

  (1) can be established by expressing both $F(b, d) - F(b, c)$ and $F(a, d) - F(a, c)$ as probabilities of events.

- Example 2. Suppose that (X, Y) has joint CDF $F_{X,Y}$ given by

$$F_{X,Y}(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}) & \text{if } x \geq 0 \text{ and } y \geq 0; \\ 0 & \text{otherwise.} \end{cases}$$

  Find $P(0 < X \leq 1 \text{ and } 2 < Y \leq 3)$.

  Sol.

$$\begin{aligned} & P(0 < X \leq 1 \text{ and } 2 < Y \leq 3) \\ &= F_{X,Y}(1, 3) - F_{X,Y}(1, 2) - F_{X,Y}(0, 3) + F_{X,Y}(0, 2) \\ &= (1 - e^{-1})(1 - e^{-3}) - (1 - e^{-1})(1 - e^{-2}) - (1 - e^0)(1 - e^{-3}) + (1 - e^0)(1 - e^{-2}) \\ &= (1 - e^{-1})(e^{-2} - e^{-3}) = e^{-2} - 2e^{-3} + e^{-4}. \end{aligned}$$

- When X and Y are discrete random variables defined on the same sample space, the joint distribution of (X, Y) can also be characterized using the joint PMF (probability mass function) of (X, Y). Let $p_{X,Y}$ be the joint PMF of (X, Y), then

$$p_{X,Y}(x, y) = P((X, Y) = (x, y))$$

  for $(x, y) \in R^2$.

- Suppose that (X, Y) is a vector of discrete random variables with joint PMF $p_{X,Y}$, then

$$P((X,Y) \in A) = \sum_{p_{X,Y}(x,y)>0 \text{ and } (x,y)\in A} p_{X,Y}(x,y)$$

for $A \in \mathcal{B}(R^2)$.

  - When $A = R^2$, we have

$$1 = \sum_{p_{X,Y}(x,y)>0} p_{X,Y}(x,y).$$

  - When $A = (-\infty, \infty) \times \{y_0\}$, we have

$$P(Y = y_0) = \sum_{p_{X,Y}(x,y)>0 \text{ and } y=y_0} p_{X,Y}(x,y) = \sum_{x:p_{X,Y}(x,y_0)>0} p_{X,Y}(x,y_0)$$

  - When $A = \{x_0\} \times (-\infty, \infty)$, we have

$$P(X = x_0) = \sum_{p_{X,Y}(x,y)>0 \text{ and } x=x_0} p_{X,Y}(x,y) = \sum_{y:p_{X,Y}(x_0,y)>0} p_{X,Y}(x_0,y)$$

- Example 3. Suppose that (X, Y) is a vector of discrete random variables with joint PMF $p_{X,Y}$, where

$$p_{X,Y}(x,y) = \begin{cases} 0.5 & \text{if } (x,y) = (1,2); \\ 0.3 & \text{if } (x,y) = (3,4); \\ 0.1 & \text{if } (x,y) = (3,6); \\ 0.1 & \text{if } (x,y) = (3,7); \\ 0 & \text{otherwise.} \end{cases}$$

Find $P((X,Y) \in (-\infty, 2] \times (-\infty, 3]) = P(X \leq 2 \text{ and } Y \leq 3)$ and the (marginal) PMF of X.

Sol. $P((X,Y) \in (-\infty, 2] \times (-\infty, 3]) = p_{X,Y}(1,2) = 0.5$. Let p_X be the PMF of X. To find p_X, note that the possible values of X are 1 and 3, $P(X = 1) = p_{X,Y}(1,2) = 0.5$ and $P(X = 3) = p_{X,Y}(3,4) + p_{X,Y}(3,6) + p_{X,Y}(3,7) = 0.3 + 0.1 + 0.1 = 0.5$, so the PMF of X is

$$p_X(x) = \begin{cases} 0.5 & \text{if } x = 1; \\ 0.5 & \text{if } x = 3; \\ 0 & \text{otherwise.} \end{cases}$$

- Suppose that (X, Y) is a vector of random variables with joint CDF $F_{X,Y}$ and

$$F_{X,Y}(s,t) = \int_{(-\infty,s]\times(-\infty,t]} f_{X,Y}(x,y)d(x,y) = \int_{-\infty}^{s}\int_{-\infty}^{t} f_{X,Y}(x,y)dydx$$

for $(s,t) \in R^2$ for some nonnegative function $f_{X,Y}$, then $f_{X,Y}$ is called the joint PDF (probability density function) of (X, Y).

- Suppose that (X,Y) has joint PDF $f_{X,Y}$, then

$$P((X,Y)\in A)=\int_A f_{X,Y}(x,y)d(x,y) \tag{2}$$

for $A\in\mathcal{B}(R^2)$.

- In (2), when $A=R^2$, we have

$$1=\int_{R^2} f_{X,Y}(x,y)d(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx.$$

- In (2), when $A=(-\infty,t]\times(-\infty,\infty)$, we have

$$P(X\le t)=\int_{-\infty}^{t}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx$$

for $t\in(\infty,\infty)$. Define

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

for $x\in(-\infty,\infty)$, then

$$\int_{-\infty}^{t} f_X(x)dx=P(X\le t)$$

for $t\in(\infty,\infty)$, so f_X is a PDF of X. Similarly, define

$$f_Y(y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

for $y\in(-\infty,\infty)$, then f_Y is a PDF of Y.

- Example 4. Suppose that (X,Y) has joint PDF $f_{X,Y}$, where

$$f_{X,Y}(x,y)=\begin{cases} ce^{-x}e^{-y} & \text{if } x>0 \text{ and } y>0;\\ 0 & \text{otherwise}\end{cases}$$

and $c>0$ is a constant.

(a) Find c.

(b) Find $P(X\le s \text{ and } Y\le t)$ for $s>0$, $t>0$.

(c) Find a PDF of X.

A sketch of solution.

(a) Solving c based on the equation $\int_{R^2} f_{X,Y}(x,y)d(x,y)=1$ gives $c=1$.

(b) For $s > 0$, $t > 0$

$$\begin{aligned} P(X \le s \text{ and } Y \le t) &= \int_{(-\infty,s]\times(-\infty,t]} f_{X,Y}(x,y)d(x,y) \\ &= (1-e^{-s})(1-e^{-t}). \end{aligned}$$

(c) Let

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

for $x \in (-\infty, \infty)$, then

$$f_X(x) = \begin{cases} e^{-x} & \text{if } x > 0; \\ 0 & \text{otherwise} \end{cases}$$

and f_X is a PDF of X.

- Example 5. Suppose that (X, Y) has joint PDF $f_{X,Y}$, where

$$f_{X,Y}(x,y) = \begin{cases} cxy & \text{if } x > 0, y > 0 \text{ and } x + y \le 1; \\ 0 & \text{otherwise} \end{cases}$$

and $c > 0$ is a constant.

(a) Find c.

(b) Find $P((X + Y) \le 0.5)$.

(c) Find a PDF of X.

Ans. (a) $c = 24$ (b) $1/16$ (c) Let $f_X(x) = 12(x^3 - 2x^2 + x)$ for $x \in (0, 1)$ and $f_X(x) = 0$ for $x \notin (0, 1)$, then f_X is a PDF of X.