5.1.2 Joint Cumulative Distributive Function (CDF)

Remember that, for a random variable X, we define the CDF as $F_X(x) = P(X \leq x)$. Now, if we have two random variables X and Y and we would like to study them jointly, we can define the **joint cumulative function** as follows:

> The **joint cumulative distribution function** of two random variables X and Y is defined as
>
> $$F_{XY}(x,y) = P(X \leq x, Y \leq y).$$

As usual, comma means "and," so we can write

$$\begin{aligned} F_{XY}(x,y) &= P(X \leq x, Y \leq y) \\ &= P\big((X \leq x) \text{ and } (Y \leq y)\big) = P\big((X \leq x) \cap (Y \leq y)\big). \end{aligned}$$

Figure 5.2 shows the region associated with $F_{XY}(x,y)$ in the two-dimensional plane. Note that the above definition of joint CDF is a general definition and is applicable to discrete, continuous, and mixed random variables. Since the joint CDF refers to the probability of an event, we must have $0 \leq F_{XY}(x,y) \leq 1$.

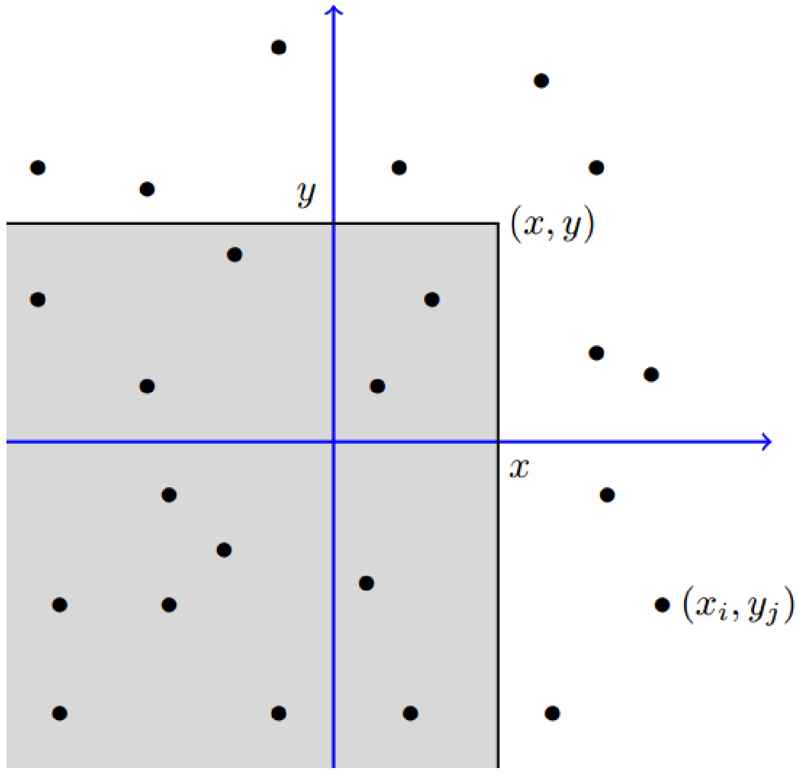

Figure 5.2: $F_{XY}(x, y)$ is the probability that (X, Y) belongs to the shaded region. The dots are the pairs (x_i, y_j) in R_{XY}.

If we know the joint CDF of X and Y, we can find the *marginal* CDFs, $F_X(x)$ and $F_Y(y)$. Specifically, for any $x \in \mathbb{R}$, we have

$$\begin{aligned} F_{XY}(x, \infty) &= P(X \leq x, Y \leq \infty) \\ &= P(X \leq x) = F_X(x). \end{aligned}$$

Here, by $F_{XY}(x, \infty)$, we mean $\lim_{y \to \infty} F_{XY}(x, y)$. Similarly, for any $y \in \mathbb{R}$, we have

$$F_Y(y) = F_{XY}(\infty, y).$$

Marginal CDFs of X and Y:

$$\begin{aligned} F_X(x) &= F_{XY}(x, \infty) = \lim_{y \to \infty} F_{XY}(x, y), && \text{for any } x, \\ F_Y(y) &= F_{XY}(\infty, y) = \lim_{x \to \infty} F_{XY}(x, y), && \text{for any } y \end{aligned} \tag{5.2}$$

Also, note that we must have

$$\begin{aligned} &F_{XY}(\infty, \infty) = 1, \\ &F_{XY}(-\infty, y) = 0, \qquad \text{for any } y, \\ &F_{XY}(x, -\infty) = 0, \qquad \text{for any } x. \end{aligned}$$

Example 5.2

Let $X \sim Bernoulli(p)$ and $Y \sim Bernoulli(q)$ be independent, where $0 < p, q < 1$. Find the joint PMF and joint CDF for X and Y.

Solution

First note that the joint range of X and Y is given by

$$R_{XY} = \{(0,0), (0,1), (1,0), (1,1)\}.$$

Since X and Y are independent, we have

$$P_{XY}(i,j) = P_X(i)P_Y(j), \qquad \text{for } i,j = 0,1.$$

Thus, we conclude

$$\begin{aligned} P_{XY}(0,0) &= P_X(0)P_Y(0) = (1-p)(1-q), \\ P_{XY}(0,1) &= P_X(0)P_Y(1) = (1-p)q, \\ P_{XY}(1,0) &= P_X(1)P_Y(0) = p(1-q), \\ P_{XY}(1,1) &= P_X(1)P_Y(1) = pq. \end{aligned}$$

Now that we have the joint PMF, we can find the joint CDF

$$F_{XY}(x,y) = P(X \leq x, Y \leq y).$$

Specifically, since $0 \leq X, Y \leq 1$, we conclude

$$\begin{aligned} &F_{XY}(x,y) = 0, \qquad \text{if } x < 0, \\ &F_{XY}(x,y) = 0, \qquad \text{if } y < 0, \\ &F_{XY}(x,y) = 1, \qquad \text{if } x \geq 1 \text{ and } y \geq 1. \end{aligned}$$

Now, for $0 \leq x < 1$ and $y \geq 1$, we have

$$\begin{aligned} F_{XY}(x,y) &= P(X \leq x, Y \leq y) \\ &= P(X = 0, y \leq 1) \\ &= P(X = 0) = 1 - p. \end{aligned}$$

Similarly, for $0 \leq y < 1$ and $x \geq 1$, we have

$$\begin{aligned} F_{XY}(x,y) &= P(X \leq x, Y \leq y) \\ &= P(X \leq 1, y = 0) \\ &= P(Y = 0) = 1 - q. \end{aligned}$$

Finally, for $0 \leq x < 1$ and $0 \leq y < 1$, we have

$$\begin{aligned} F_{XY}(x, y) &= P(X \leq x, Y \leq y) \\ &= P(X = 0, y = 0) \\ &= P(X = 0)P(Y = 0) = (1 - p)(1 - q). \end{aligned}$$

Figure 5.3 shows the values of $F_{XY}(x, y)$ in different regions of the two-dimensional plane. Note that, in general, we actually need a three-dimensional graph to show a joint CDF of two random variables, i.e., we need three axes: x, y, and $z = F_{XY}(x, y)$. However, because the random variables of this example are simple, and can take only two values, a two-dimensional figure suffices.

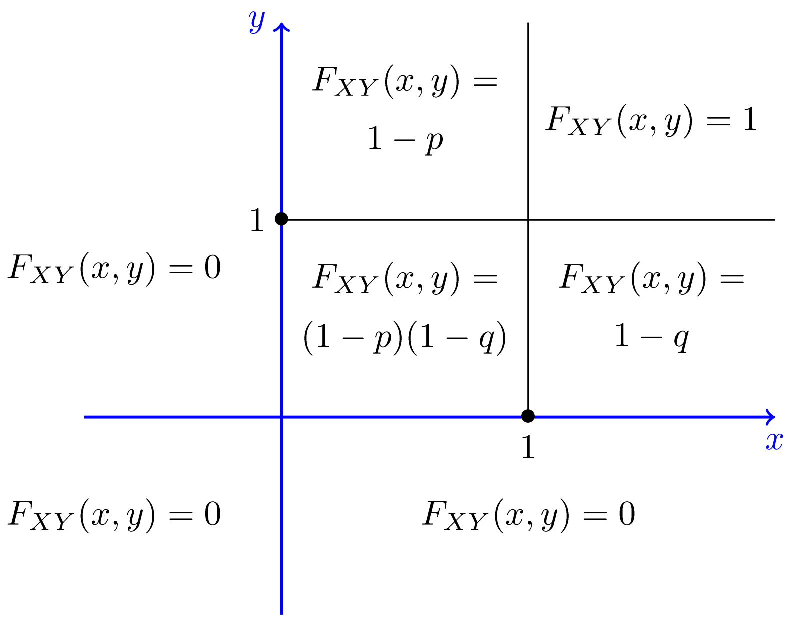

Figure 5.3 Joint CDF for X and Y in Example 5.2

Here is a useful lemma:

Lemma 5.1
For two random variables X and Y, and real numbers $x_1 \leq x_2$, $y_1 \leq y_2$, we have

$$\begin{aligned} & P(x_1 < X \leq x_2, \ \ y_1 < Y \leq y_2) = \\ & \qquad F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1). \end{aligned}$$

To see why the above formula is true, you can look at the region associated with $F_{XY}(x, y)$ (as shown in Figure 5.2) for each of the pairs $(x_2, y_2), (x_1, y_2), (x_2, y_1), (x_1, y_1)$. You can see, as we subtract and add regions, the part that is left is the region $\{x_1 < X \leq x_2, \ \ y_1 < Y \leq y_2\}$.