Jointly Distributed Random Variables

Two random variables [Ross S6.1]

So far, we only considered the distribution of a single random variable.

Say we want the probability of an event involving 2 or more random variables:

i) $P[X < 3, Y > 7]$
ii) $P[X < Y]$
iii) $P[X^2 + Y^2 < 10]$
iv) $P[XY = 3]$

For this, we need the **joint cumulative distribution function** (joint CDF):

$$F_{XY}(a, b) = P[X \le a, Y \le b]$$

All probability statements involving X and Y can be found from $F_{XY}(a, b)$.

Example 22.1: For $a_1 < a_2$ and $b_1 < b_2$, show that

$$\begin{aligned} &P[a_1 < X \le a_2, b_1 < Y \le b_2] \\ &\quad = F_{XY}(a_2, b_2) + F_{XY}(a_1, b_1) - F_{XY}(a_1, b_2) - F_{XY}(a_2, b_1) \end{aligned}$$

Solution:

$$\begin{aligned} F_{XY}(a_2, b_2) &= P[X \le a_1, Y \le b_1] + P[a_1 < X \le a_2, Y \le b_1] \\ &\quad + P[X \le a_1, b_1 < Y \le b_2] + P[a_1 < X \le a_2, b_1 < Y \le b_2] \\ F_{XY}(a_1, b_2) &= P[X \le a_1, Y \le b_1] + P[X \le a_1, b_1 < Y \le b_2] \\ F_{XY}(a_2, b_1) &= P[X \le a_1, Y \le b_1] + P[a_1 < X \le a_2, Y \le b_1] \\ F_{XY}(a_1, b_1) &= P[X \le a_1, Y \le b_1] \end{aligned}$$

Combining we get the desired result.

Discrete Case:

Say X and Y are both discrete:

- X takes values in $\mathcal{X} = \{x_1, x_2, \ldots\}$,
- Y takes values in $\mathcal{Y} = \{y_1, y_2, \ldots\}$.

We define the **joint probability mass function** (joint pmf):

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

Then
$$\begin{aligned} p_X(x) &= P[X = x] \\ &= P[\cup_j \{X = x, Y = y_j\}] \\ &= \sum_j P[X = x, Y = y_j] \\ &= \sum_j p_{XY}(x, y_j) \end{aligned}$$

Likewise $\quad p_Y(y) = \sum_i p_{XY}(x_i, y)$

Note: $p_X(x)$ is called the X **marginal** of $p_{XY}(x, y)$. This process is called **marginalization**.

Note: This is because if we list $p_{XY}(x_i, y_j)$ in a table on a page, then the sum over j is summing the ith row of the table, and writing each sum in the right margin of the page.

Also
$$\begin{aligned} 1 &= P[X \in \mathcal{X}, Y \in \mathcal{Y}] \\ &= P[\cup_{i,j} \{X = x_i, Y = y_j\}] \\ &= \sum_{i,j} P[X = x_i, Y = y_j] \\ &= \sum_{i,j} p_{XY}(x_i, y_j) \end{aligned}$$

So joint pmf must sum to 1.

Example 22.2: An urn contains 3 red, 4 white and 5 blue balls. 3 balls are picked at random. Let X = # red balls, Y = # white balls.

Find $p_{XY}(i, j)$.

Solution: There are $\binom{12}{3}$ ways of picking 3 of 12 balls.

If $X = i$ and $Y = j$, then # blue balls is $3 - i - j$.

There are:

- $\binom{3}{i}$ ways of picking i red balls from 3 red balls,
- $\binom{4}{j}$ ways of picking j white balls from 4 white balls,
- $\binom{5}{3-i-j}$ ways of picking $3 - i - j$ blue balls from 5 blue

Therefore $\quad p_{XY}(i, j) = \dfrac{\binom{3}{i}\binom{4}{j}\binom{5}{3-i-j}}{\binom{12}{3}}$

Continuous Case:

X and Y are jointly continuous random variables if there exists a non-negative $f_{XY}(x, y)$ such that for every $C \subset \mathbb{R}^2$:

$$P[(X, Y) \in C] = \iint_C f_{XY}(x, y)\,dxdy$$

$f_{XY}(x, y)$ is called the **joint probability density function** (joint pdf).

Since $P[X \in A, Y \in B] = P[(X, Y) \in \underbrace{A \times B}_{C}]$, then

$$\begin{aligned} P[X \in A, Y \in B] &= \iint_{A \times B} f_{XY}(x, y)\,dxdy \\ &= \int_B \int_A f_{XY}(x, y)\,dxdy \end{aligned}$$

Also,

$$\begin{aligned} F_{XY}(a, b) &= P[X \le a,\ Y \le b] \\ &= P[X \in (-\infty, a],\ Y \in (-\infty, b]\ ] \\ &= \int_{-\infty}^{b} \int_{-\infty}^{a} f_{XY}(x, y)\ dxdy \end{aligned} \tag{22.1}$$

Taking partial derivatives with respect to a and b in (22.1)

$$f_{XY}(a, b) = \frac{\partial^2}{\partial a \partial b} F_{XY}(a, b)$$

Also

$$\begin{aligned} \int_A f_X(x)dx &= P[X \in A] \\ &= P[X \in A, Y \in (-\infty, \infty)] \\ &= \int_A \int_{-\infty}^{\infty} f_{XY}(x, y)\ dydx \end{aligned}$$

So $\quad f_X(x) = \displaystyle\int_{-\infty}^{\infty} f_{XY}(x, y)dy \qquad$ [marginalization]

Likewise $\quad f_Y(y) = \displaystyle\int_{-\infty}^{\infty} f_{XY}(x, y)dx$

Also,

$$\begin{aligned} 1 &= P[X \in (-\infty, \infty), Y \in (-\infty, \infty)] \\ &= \iint_{\mathbb{R}^2} f_{XY}(x, y)dxdy \end{aligned}$$

So for a joint pdf, volume under the curve is 1.