6.1.1 Joint Distributions and Independence

For three or more random variables, the joint PDF, joint PMF, and joint CDF are defined in a similar way to what we have already seen for the case of two random variables. Let X_1, X_2, $\cdots$, X_n be n discrete random variables. The joint PMF of X_1, X_2, $\cdots$, X_n is defined as

$$P_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)=P(X_1=x_1,X_2=x_2,\ldots,X_n=x_n).$$

For n jointly continuous random variables X_1, X_2, $\cdots$, X_n, the joint PDF is defined to be the function $f_{X_1X_2\ldots X_n}(x_1,x_2,\ldots,x_n)$ such that the probability of any set $A\subset\mathbb{R}^n$ is given by the integral of the PDF over the set A. In particular, for a set $A\in\mathbb{R}^n$, we can write

$$P\Big((X_1,X_2,\cdots,X_n)\in A\Big)=\int\cdots\underset{A}{\int}\cdots\int f_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)dx_1dx_2\cdots dx_n.$$

The marginal PDF of X_i can be obtained by integrating all other X_j's. For example,

$$f_{X_1}(x_1)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_{X_1X_2\ldots X_n}(x_1,x_2,\ldots,x_n)dx_2\cdots dx_n.$$

The joint CDF of n random variables X_1, X_2,...,X_n is defined as

$$F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)=P(X_1\leq x_1,X_2\leq x_2,\ldots,X_n\leq x_n).$$

Example 6.1

Let X,Y and Z be three jointly continuous random variables with joint PDF

$$f_{XYZ}(x,y,z)=\begin{cases} c(x+2y+3z) & 0\leq x,y,z\leq 1\\ \\ 0 & \text{otherwise}\end{cases}$$

1. Find the constant c.
2. Find the marginal PDF of X.

Solution

1.

$$
\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XYZ}(x,y,z)dxdydz \\
&= \int_0^1\int_0^1\int_0^1 c(x+2y+3z) \;\; dxdydz \\
&= \int_0^1\int_0^1 c\left(\frac{1}{2}+2y+3z\right) \;\; dydz \\
&= \int_0^1 c\left(\frac{3}{2}+3z\right) \;\; dz \\
&= 3c.
\end{aligned}
$$

Thus, $c=\frac{1}{3}$.

2. To find the marginal PDF of X, we note that $R_X=[0,1]$. For $0 \leq x \leq 1$, we can write

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XYZ}(x,y,z)dydz \\
&= \int_0^1\int_0^1 \frac{1}{3}(x+2y+3z) \;\; dydz \\
&= \int_0^1 \frac{1}{3}(x+1+3z) \;\; dz \\
&= \frac{1}{3}\left(x+\frac{5}{2}\right).
\end{aligned}
$$

Thus,

$$
f_X(x) = \begin{cases} \frac{1}{3}\left(x+\frac{5}{2}\right) & 0 \leq x \leq 1 \\ \\ 0 & \text{otherwise} \end{cases}
$$

Independence: The idea of independence is exactly the same as what we have seen before. We restate it here in terms of the joint PMF, joint PDF, and joint CDF. Random variables X_1, X_2, ... , X_n are independent, if for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$,

$$
F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n).
$$

Equivalently, if X_1, X_2, ..., X_n are discrete, then they are independent if for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, we have

$$
P_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P_{X_1}(x_1)P_{X_2}(x_2)\cdots P_{X_n}(x_n).
$$

If X_1, X_2, ..., X_n are continuous, then they are independent if for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, we have

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n).$$

If random variables X_1, X_2, ..., X_n are independent, then we have

$$E[X_1 X_2 \cdots X_n] = E[X_1]E[X_2] \cdots E[X_n].$$

In some situations we are dealing with random variables that are independent and are also identically distributed, i.e, they have the same CDFs. It is usually easier to deal with such random variables, since independence and being identically distributed often simplify the analysis. We will see examples of such analyses shortly.

Definition 6.1. Random variables X_1, X_2, ..., X_n are said to be **independent and identically distributed (i.i.d.)** if they are *independent*, and they have the *same marginal distributions*:

$$F_{X_1}(x) = F_{X_2}(x) = \ldots = F_{X_n}(x), \text{ for all } x \in \mathbb{R}.$$

For example, if random variables X_1, X_2, ..., X_n are i.i.d., they will have the same means and variances, so we can write

$$\begin{aligned} E[X_1 X_2 \cdots X_n] &= E[X_1]E[X_2] \cdots E[X_n] && \text{(because the } X_i\text{'s are indepenednt)} \\ &= E[X_1]E[X_1] \cdots E[X_1] && \text{(because the } X_i\text{'s are identically distributed)} \\ &= E[X_1]^n. \end{aligned}$$