

PROBABILITY

QSNKE EDITION

1. Joint Distribution Functions.

For any two random variables X and Y , the joint cumulative probability distribution function of X and Y by

$$F(a, b) = P\{X \leq a, Y \leq b\} \quad -\infty < a, b < \infty$$

The distribution of X can be obtained from the joint distribution of X and Y as follows:

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\{X \leq a, Y \leq b\} \\ &= P\left(\lim_{b \rightarrow \infty} \{X \leq a, Y \leq b\}\right) \\ &= \lim_{b \rightarrow \infty} P\{X \leq a, Y \leq b\} \\ &= \lim_{b \rightarrow \infty} F(a, b) \\ &= F(a, \infty) \end{aligned}$$

and same as the distribution of X . The distribution functions F_X and F_Y are sometimes referred to as the marginal distributions of X and Y .

If we want to compute the joint probability that X is greater than a and Y is greater than b . This could be done as follows:

$$\begin{aligned} P\{X > a, Y > b\} &= 1 - P(\{X > a, Y > b\}^c) \\ &= 1 - P(\{X > a\}^c \cup \{Y > b\}^c) \\ &= 1 - P(\{X \leq a\} \cup \{Y \leq b\}) \\ &= 1 - [P\{X \leq a\} + P\{Y \leq b\} - P\{X \leq a, Y \leq b\}] \\ &= 1 - F_X(a) - F_Y(b) + F(a, b) \end{aligned}$$

and the above equation is a special case of the following equation, whose verification is left as an exercise:

$$P\{a_1 < X \leq a_2, b_1 < Y \leq b_2\} = F(a_2, b_2) + F(a_1, b_1) - F(a_1, b_2) - F(a_2, b_1)$$

whenever $a_1 < a_2, b_1 < b_2$

In the case when X and Y are both discrete random variables, it is convenient to define the joint probability mass function of X and Y by

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

2. Independent Random Variables.

The random variables X and Y are said to be independent if, for any two sets of real numbers A and B ,

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

In other words, X and Y are independent if, for all A and B , the events $E_A = \{X \in A\}$ and $E_B = \{Y \in B\}$ are independent.

It can be shown by using the three axioms of probability that will follow if and only if, for all a, b ,

$$P\{X \leq a, Y \leq b\} = P\{X \leq a\}P\{Y \leq b\}$$

Hence, in terms of the joint distribution function F of X and Y , X and Y are independent if

$$F(a, b) = F_X(a)F_Y(b) \quad \forall a, b$$

When X and Y are discrete random variables, the condition of independence is equivalent to

$$p(x, y) = p_X(x)p_Y(y) \quad \forall x, y$$

Example. Buffon's needle problem

A table is ruled with equidistant parallel lines a distance D apart. A needle of length L , where $L \leq D$, is randomly thrown on the table. What is the probability that the needle will intersect one of the lines (the other possibility being that the needle will be completely contained in the strip between two lines)?

Proposition. The continuous (discrete) random variables X and Y are independent if and only if their probability density (mass) function can be expressed as

$$f_{X,Y}(x, y) = h(x)g(y) \quad -\infty < x < \infty, \quad -\infty < y < \infty$$

Proof. Let us give the proof in the continuous case. First, note that independence implies that the joint density is the product of the marginal densities of X and Y , so the preceding factorization will hold when the random variables are independent. Now, suppose that

$$F_{X,Y}(x, y) = h(x)g(y)$$

Then

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} h(x) dx \int_{-\infty}^{\infty} g(y) dy \\ &= C_1 C_2 \end{aligned}$$

where $C_1 = \int_{-\infty}^{\infty} h(x) dx$ and $C_2 = \int_{-\infty}^{\infty} g(y) dy$. Also,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = C_2 h(x)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = C_1 g(y)$$

Since $C_1 C_2 = 1$, it follows that

$$F_{X,Y}(x, y) = f_X(x)f_Y(y)$$

■

3. Sums of Independent Random Variables.

Suppose X and Y are independent, continuous random variables having probability density functions F_X and f_Y . The cumulative distribution function of $X + Y$ is obtained as follows:

$$\begin{aligned}
 F_{X+Y}(a) &= P\{X + Y \leq a\} \\
 &= \iint_{x+y \leq a} f_X(x)f_Y(y)dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x)f_Y(y)dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x)dx f_Y(y)dy \\
 &= \int_{-\infty}^{\infty} F_X(a - y)f_Y(y)dy
 \end{aligned}$$

The cumulative distribution function F_{X+Y} is called the convolution of the distribution F_X and F_Y . By differentiating above equation, we find that the probability density function F_{X+Y} of $X + Y$ is given by

$$\begin{aligned}
 f_{X+Y}(a) &= \frac{d}{da} \int_{-\infty}^{\infty} F_X(a - y)f_Y(y)dy \\
 &= \int_{-\infty}^{\infty} \frac{d}{da} F_X(a - y)f_Y(y)dy \\
 &= \int_{-\infty}^{\infty} f_X(a - y)f_Y(y)dy
 \end{aligned}$$