

# PROBABILITY

QSSNAKE 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:

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

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,

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

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

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

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

we denote the equation as  $(\star)$

§ **3-1 Identically Distributed Uniform Random Variables.** If  $X$  and  $T$  are independent random variables, both uniformly distributed in  $(0, 1)$ , calculate the probability density of  $X + Y$ .

From the equation above, since

$$f_X(a) = f_Y(a) = \begin{cases} 1 & 0 < a < 1 \\ 0 & \text{otherwise} \end{cases}$$

we obtain

$$f_{X+Y}(a) = \int_0^1 f_X(a-y)dy$$

For  $0 \leq a \leq 1$ , this yields

$$f_{X+Y}(a) = \int_0^a dy = a$$

For  $1 < a < 2$ , we get

$$f_{X+Y}(a) = \int_{a-1}^1 dy = 2 - a$$

Hence,

$$f_{X+Y}(a) = \begin{cases} a & 0 \leq a \leq 1 \\ 2 - a & 1 < a < 2 \\ 0 & \text{otherwise} \end{cases}$$

Because of the shape of its density function, the random variable  $X + Y$  is said to have a triangular distribution.

**3.2 Gamma Random Variables.** A gamma random variable has a density of the form

$$f(y) = \frac{\lambda e^{-\lambda y} (\lambda y)^{t-1}}{\Gamma(t)} \quad 0 < y < \infty$$

**Prop.** If  $X$  and  $Y$  are independent gamma random variables with respective parameters  $(s, \lambda)$  and  $(t, \lambda)$ , then  $X + Y$  is a gamma random variable with parameters  $(s + t, \lambda)$

*Proof.* Using equation  $(\star)$ , we obtain

$$\begin{aligned}
 f_{X+Y}(a) &= \frac{1}{\Gamma(s)\Gamma(t)} \int_0^a \lambda e^{-\lambda(a-y)} [\lambda(a-y)]^{s-1} \lambda e^{-\lambda y} (\lambda y)^{t-1} dy \\
 &= K e^{-\lambda a} \int_0^a (a-y)^{s-1} y^{t-1} dy \\
 &= K e^{-\lambda a} a^{s+t-1} \int_0^1 (1-x)^{s-1} x^{t-1} dx \text{ by letting } x = \frac{y}{a} \\
 &= C e^{-\lambda a} a^{s+t-1}
 \end{aligned}$$

where  $C$  is a constant that does not depend on  $a$ . But, as the preceding is a density function and thus must integrate to 1, the value of  $C$  is determined, and we have

$$f_{X+Y}(a) = \frac{\lambda e^{-\lambda a} (\lambda a)^{s+t-1}}{\Gamma(s+t)}$$

Hence, the result is proved. ■

**Normal Random Variables.** We can also use equation  $(\star)$  to prove the following important result about normal random variables.

**Prop.** If  $X_i$ ,  $i = 1, \dots, n$ , are independent random variables that are normally distributed with respective parameters  $\mu_i, \sigma_i^2$ ,  $i = 1, \dots, n$ , then  $\sum_{i=1}^n X_i$  is normally distributed with parameters  $\sum_{i=1}^n \mu_i$  and  $\sum_{i=1}^n \sigma_i^2$

*Proof.* To begin, let  $X$  and  $Y$  be independent normal random variables with  $X$  having mean 0 and variance  $\sigma^2$  and  $Y$  having mean 0 and variance 1. We will determine the density function of  $X+Y$  by utilizing equation  $(\star)$ . Now, with

$$c = \frac{1}{2\sigma^2} + \frac{1}{2} = \frac{1 + \sigma^2}{2\sigma^2}$$

we have

$$\begin{aligned}
 f_X(a-y)f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(a-y)^2}{2\sigma^2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} \\
 &= \frac{1}{2\pi\sigma} - \frac{a^2}{2\sigma^2} \exp\left\{-\frac{a^2}{2\sigma^2}\right\} \exp\left\{-c\left(y^2 - 2y\frac{a}{1+\sigma^2}\right)\right\}
 \end{aligned}$$

6

Hence, from equation  $(\star)$

$$\begin{aligned}
 f_{X+Y}(a) &= \frac{1}{2\pi\sigma} \exp\left\{-\frac{a^2}{2\sigma^2}\right\} \exp\left\{\frac{a^2}{2\sigma^2(1+\sigma^2)}\right\} \\
 &\quad \times \int_{-\infty}^{\infty} \exp\left\{-c\left(y - \frac{a}{1+\sigma^2}\right)\right\} \\
 &= \frac{1}{2\pi\sigma} \exp\left\{-\frac{a^2}{2(1+\sigma^2)}\right\} \int_{-\infty}^{\infty} \exp\{-cx^2\} dx \\
 &= C \exp\left\{-\frac{a^2}{2(1+\sigma^2)}\right\}
 \end{aligned}$$

■

where  $C$  does not depend on  $a$ . But this implies that  $X + Y$  is normal with mean 0 and variance  $1 + \sigma^2$ .

Now, suppose that  $X_1$  and  $X_2$  are independent normal random variables with  $X_i$  having mean  $\mu_i$  and variance  $\sigma_i^2$ ,  $i = 1, 2$ . Then

$$X_1 + X_2 = \sigma_2 \left( \frac{X_1 - \mu_1}{\sigma_2} + \frac{X_2 - \mu_2}{\sigma_2} \right) + \mu_1 + \mu_2$$

But since  $(X_1 - \mu_1)/\sigma_2$  is normal with mean 0 and variance  $\sigma_1^2/\sigma_2^2$ , and  $(X_2 - \mu_2)/\sigma_2$  is normal with mean 0 and variance 1, it follows from our previous result that  $(X_1 - \mu_1)/\sigma_2 + (X_2 - \mu_2)/\sigma_2$  is normal with mean 0 and variance  $1 + \sigma_1^2/\sigma_2^2$ , implying that  $X_1 + X_2$  is normal with mean  $\mu_1 + \mu_2$  and variance  $\sigma_2^2(1 + \sigma_1^2/\sigma_2^2) = \sigma_1^2 + \sigma_2^2$ .

Thus, Proposition is established when  $n = 2$ . The general case now follows by induction. That is, assume that Proposition is true when there are  $n - 1$  random variables. Now consider the case of  $n$ , and write

$$\sum_{i=1}^n X_i = \sum_{i=1}^{n-1} X_i + X_n$$

By the induction hypothesis,  $\sum_{i=1}^{n-1} X_i$  is normal with mean  $\sum_{i=1}^{n-1} \mu_i$  and variance  $\sum_{i=1}^{n-1} \sigma_i^2$ . Therefore, by the result for  $n = 2$ ,  $\sum_{i=1}^n X_i$  is normal with mean  $\sum_{i=1}^n \mu_i$  and variance  $\sum_{i=1}^n \sigma_i^2$ .

**Poisson and Binomial Random Variables.** Rather than attempt to derive a general expression for the distribution of  $X + Y$  in the discrete case, we shall consider some examples.

**Example.** Sums of independent Poisson random variables

If  $X$  and  $Y$  are independent Poisson random variables with respective parameters  $\lambda_1$  and  $\lambda_2$ , compute the distribution of  $X + Y$ .

*Solution.* Because the event  $\{X + Y = n\}$  may be written as the union of the disjoint events  $\{X = k, Y = n - k\}, 0 \leq k \leq n$  may be written as the union of the disjoint events  $\{X = k, Y = n - k\}, 0 \leq k \leq n$ , we have

$$\begin{aligned}
 P\{X + Y = n\} &= \sum_{k=0}^n P\{X = k, Y = n - k\} \\
 &= \sum_{k=0}^n P\{X = k\}P\{Y = n - k\} \\
 &= \sum_{k=0}^n e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \\
 &= e^{-(\lambda_1 + \lambda_2)} \sum_{k=0}^n \frac{\lambda_1^k \lambda_2^{n-k}}{k!(n-k)!} \\
 &= \frac{e^{-\lambda_1 + \lambda_2}}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} \\
 &= \frac{e^{-(\lambda_1 + \lambda_2)}}{n!} (\lambda_1 + \lambda_2)^n
 \end{aligned}$$

Thus,  $X + Y$  has a Poisson distribution with parameter  $\lambda_1 + \lambda_2$  ■

**Example.** Sums of independent binomial random variables

Let  $X$  and  $Y$  be independent binomial random variables with respective parameters  $(n, p)$  and  $(m, p)$ . Calculate the distribution of  $X + Y$ .



*Solution.*

$$\begin{aligned}
 P\{X + Y = k\} &= \sum_{i=0}^n P\{X = i, Y = k - i\} \\
 &= \sum_{i=0}^n P\{X = i\}P\{Y = k - i\} \\
 &= \sum_{i=0}^n \binom{n}{i} p^i q^{n-i} \binom{m}{k-i} p^{k-i} q^{m-k+i}
 \end{aligned}$$

where  $q = 1 - p$  and where  $\binom{r}{j} = 0$  when  $j < 0$ . Thus,

$$P\{X + Y = k\} = p^k q^{n+m-k} \sum_0^n \binom{n}{i} \binom{m}{k-i}$$

and the conclusion follows upon application of the combinatorial identity

$$\binom{n+m}{k} = \sum_{i=0}^n \binom{n}{i} \binom{m}{k-i}$$

■

**4. Conditional Distributions: Discrete Case.** For any two events  $E$  and  $F$ , the conditional probability of  $E$  given  $F$  is defined, provided that  $P(F) > 0$ , by

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Hence, if  $X$  and  $Y$  are discrete random variables, it is natural to define the conditional probability mass function of  $X$  given that  $Y = y$ , by

$$\begin{aligned} p_{X|Y}(x|y) &= P\{X = x \mid Y = y\} \\ &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\ &= \frac{p(x, y)}{p_Y(y)} \end{aligned}$$

for all values of  $y$  such that  $p_Y(y) > 0$ . Similarly, the conditional probability distribution function of  $X$  given that  $Y = y$  is defined, for all  $y$  such that  $p_Y(y) > 0$ , by

$$\begin{aligned} F_{X|Y}(x|y) &= P\{X \leq x \mid Y = y\} \\ &= \sum_{a \leq x} p_{X|Y}(a|y) \end{aligned}$$

In other words, the definitions are exactly the same as in unconditional case, except that every thing is now conditional on the event that  $Y = y$ . If  $X$  is independent of  $Y$ , then the conditional mass function and the distribution function are the same as the respective unconditional ones. This follows because if  $X$  is independent of  $Y$ , then

$$\begin{aligned} P_{X|Y}(x|y) &= P\{X = x \mid Y = y\} \\ &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\ &= \frac{P\{X = x\}P\{Y = y\}}{P\{Y = y\}} \\ &= P\{X = x\} \end{aligned}$$

**Exmample.** If  $X$  and  $Y$  are independent Poisson random variables with respective parameters  $\lambda_1$  and  $\lambda_2$ , calculate the conditional distribution of  $X$  given that  $X + Y = n$ .

*Solution.* We calculate the conditional probability mass function of  $X$  given that  $X + Y = n$  as follows:

$$\begin{aligned} P\{X = k \mid X + Y = n\} &= \frac{P\{X = k, X + Y = n\}}{P\{X + Y = n\}} \\ &= \frac{P\{X = k, Y = n - k\}}{P\{X + Y = n\}} \\ &= \frac{P\{X = k\}P\{Y = n - k\}}{P\{X + Y = n\}} \end{aligned}$$

where the last equality follows from the assumed independence of  $X$  and  $Y$ . Recalling that  $X + Y$  has a Poisson distribution with parameter  $\lambda_1 + \lambda_2$ , we see that the preceding equals

$$\begin{aligned} P\{X = k \mid X + Y = n\} &= \frac{e^{-\lambda_1} \lambda_1^k}{k!} \frac{e^{-\lambda_2} \lambda_2^{n-k}}{(n-k)!} \left[ \frac{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^n}{n!} \right]^{-1} \\ &= \frac{n!}{(n-k)!k!} \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n} \\ &= \binom{n}{k} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k} \end{aligned}$$

In other words, the conditional distribution of  $X$  given that  $X + Y = n$  is the binomial distribution with parameters  $n$  and  $\lambda_1/(\lambda_1 + \lambda_2)$ . ■

**5. Conditional Distributions: Continuous Case.** If  $X$  and  $Y$  have a joint probability density function  $f(x, y)$ , then the conditional probability density function of  $X$  given that  $Y = y$  is defined, for all values of  $y$  such that  $f_Y(y) > 0$ , by

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

To motivate this definition, multiply the left-hand side by  $dx$  and the right-hand side by  $(dxdy)/dy$  to obtain

$$\begin{aligned} f_{X|Y}(x|y)dx &= \frac{f(x, y)dxdy}{f_Y(y)dy} \\ &\approx \frac{P\{x \leq X \leq x + dx, y \leq Y \leq y + dy\}}{P\{y \leq Y \leq y + dy\}} \\ &= P\{x \leq X \leq x + dx \mid y \leq Y \leq y + dy\} \end{aligned}$$

The use of conditional densities allows us to define conditional probabilities of events associated with one random variable when we are given the value of a second random variable. That is, if  $X$  and  $Y$  are jointly continuous, then, for any set  $A$ ,

$$P\{X \in A \mid Y = y\} = \int_A f_{X|Y}(x|y)dx$$

In particular, by letting  $A = (-\infty, a)$  we can define the conditional cumulative distribution function of  $X$  given that  $Y = y$  by

$$F_{X|Y}(a|y) \equiv P\{X \leq a \mid Y = y\} = \int_{-\infty}^a f_{X|Y}(x|y)dx$$

If  $X$  and  $Y$  are independent continuous random variables, the conditional density of  $X$  given that  $Y = y$  is just the unconditional density of  $X$ . This is so because, in the independent case,

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x)$$

**Example.**

The joint density of  $X$  and  $Y$  is given by

$$f(x, y) = \begin{cases} \frac{12}{5}x(2 - x - y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the conditional density of  $X$  given that  $Y = y$ , where  $0 < y < 1$ .

*Solution.* For  $0 < x < 1, 0 < y < 1$ , we have

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x, y)}{f_Y(y)} \\ &= \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx} \\ &= \frac{x(2 - x - y)}{\int_0^1 x(2 - x - y) dx} \\ &= \frac{x(2 - x - y)}{\frac{2}{3} - y/2} \\ &= \frac{6x(2 - x - y)}{4 - 3y} \end{aligned}$$

■

**7. Joint Probability Distribution of Functions of Random Variables.** Let  $X_1$  and  $X_2$  be jointly continuous random variables with joint probability density function  $f_{X_1, X_2}$ . It is sometimes necessary to obtain the joint distribution of the random variables  $Y_1$  and  $Y_2$ , which arise as functions of  $X_1$  and  $X_2$ . Specifically, suppose that  $Y_1 = g_1(X_1, X_2)$  and  $Y_2 = g_2(X_1, X_2)$  for some functions  $g_1$  and  $g_2$ .

Assume that the function  $g_1$  and  $g_2$  satisfy the following conditions:

- (1) The equations  $y_1 = g_1(x_1, x_2)$  and  $y_2 = g_2(x_1, x_2)$  can be uniquely solved for  $x_1$  and  $x_2$  in terms of  $y_1$  and  $y_2$ , with solutions given by, say,  $x_1 = h_1(y_1, y_2)$ ,  $x_2 = h_2(y_1, y_2)$
- (2) The functions  $g_1$  and  $g_2$  have continuous partial derivatives at all points  $(x_1, x_2)$  and are such that the  $2 \times 2$  determinant

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \frac{\partial g_1}{\partial x_1} \frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2} \frac{\partial g_2}{\partial x_1} \neq 0$$

at all points  $(x_1, x_2)$

Under these two conditions, it can be shown that the random variables  $Y_1$  and  $Y_2$  are jointly continuous with joint density function given by

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1, X_2}(x_1, x_2) |J(x_1, x_2)|^{-1}$$

where  $x_1 = h_1(y_1, y_2)$ ,  $x_2 = h_2(y_1, y_2)$ .

A proof of equation above would proceed along the following lines:

$$P\{Y_1 \leq y_1, Y_2 \leq y_2\} = \iint_{\substack{(x_1, x_2) \\ g_1(x_1, x_2) \leq y_1 \\ g_2(x_1, x_2) \leq y_2}} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2$$