

4

RANDOM VARIABLES

7. Discrete and Continuous Random Variables. Distribution Functions

Given a sample space Ω , by a *random variable* we mean a numerical function $\xi = \xi(\omega)$ whose value depends on the elementary events $\omega \in \Omega$. Let $\mathbf{P}\{x' < \xi < x''\}$ be the probability of the event $\{x' < \xi < x''\}$, i.e., the probability that ξ takes a value in the interval $x' < x < x''$. Then knowledge of $\mathbf{P}\{x' < \xi < x''\}$ for all x' and x'' ($x' < x''$) is said to specify the *probability distribution* of the random variable ξ .

A random variable $\xi = \xi(\omega)$ is said to be *discrete* (or to have a *discrete distribution*) if ξ takes only a finite or countably infinite number of distinct values x , with corresponding probabilities

$$P_{\xi}(x) = \mathbf{P}\{\xi = x\},$$

$$\sum_{-\infty}^{\infty} P_{\xi}(x) = 1,$$

where the summation is over all the values of x taken by ξ . For such random variables,

$$\mathbf{P}\{x' < \xi < x''\} = \sum_{x'}^{x''} P_{\xi}(x), \quad (4.1)$$

where the summation is over the finite or countably infinite number of values of x which ξ can take in the interval $x' < x < x''$.

A random variable $\xi = \xi(\omega)$ is said to be *continuous* (or to have a *continuous distribution*) if

$$\mathbf{P}\{x' < \xi < x''\} = \int_{x'}^{x''} p_{\xi}(x) dx, \quad (4.2)$$

where $p_{\xi}(x)$ is a nonnegative integrable function, called the *probability density* of the random variable ξ , with unit integral

$$\int_{-\infty}^{\infty} p_{\xi}(x) dx = 1.$$

Clearly, if ξ is a continuous random variable, then

$$\mathbf{P}\{\xi = x\} = 0$$

for any given value x , while¹

$$\mathbf{P}\{\xi \in dx\} \sim p_{\xi}(x) dx$$

for every x with a neighborhood in which the probability density $p_{\xi}(x)$ is continuous. Here $\mathbf{P}\{\xi \in dx\}$ is the probability of the event $\{\xi \in dx\}$, consisting of ξ taking any value in an infinitesimal interval dx centered at the point x .

The function

$$\Phi_{\xi}(x) = \mathbf{P}\{\xi \leq x\}, \quad -\infty < x < \infty$$

is called the *distribution function* of the random variable ξ . If ξ is a discrete random variable, $\Phi_{\xi}(x)$ is the step function

$$\Phi_{\xi}(x) = \sum_{-\infty}^x p_{\xi}(x),$$

taking a finite or countably infinite number of distinct values [the graph of such a function is shown in Figure 4(a)]. If ξ is a continuous random variable, $\Phi_{\xi}(x)$ is the continuous function

$$\Phi_{\xi}(x) = \int_{-\infty}^x p_{\xi}(x) dx \quad (4.3)$$

[the graph of such a function is shown in Figure 4(b)].² Clearly,

$$\mathbf{P}\{x' < \xi \leq x''\} = \Phi_{\xi}(x'') - \Phi_{\xi}(x') \quad (4.4)$$

for any random variable ξ .

Now consider two random variables ξ_1 and ξ_2 , or equivalently, the random point or vector $\xi = (\xi_1, \xi_2)$. First suppose ξ_1 and ξ_2 are discrete. Then ξ_1 and ξ_2 have a *joint probability distribution*, characterized by the probabilities

$$p_{\xi_1, \xi_2}(x_1, x_2) = \mathbf{P}\{\xi_1 = x_1, \xi_2 = x_2\}, \quad (4.5)$$

where x_1 and x_2 range over all possible values of the corresponding random

¹ The symbol $\in$ means "belongs to" or "is contained in."

² By a well-known theorem on differentiation, $d\Phi_{\xi}(x)/dx = p_{\xi}(x)$ almost everywhere. See e.g., E. C. Titchmarsh, *The Theory of Functions*, second edition, Oxford University Press, London (1939), p. 362.

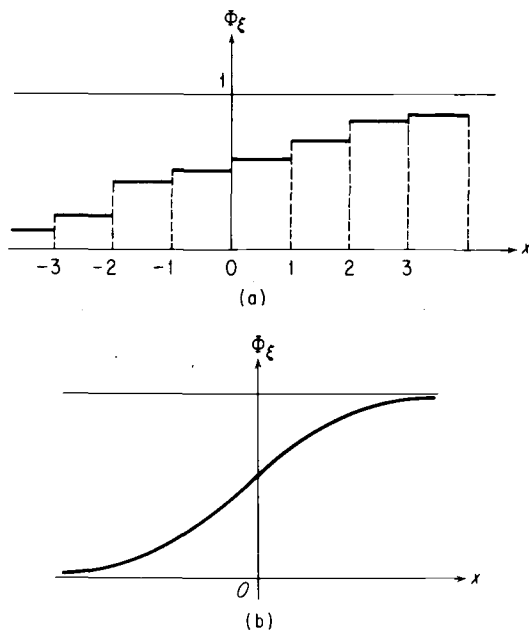


FIGURE 4. (a) A typical distribution function of a discrete random variable taking only the integral values $\dots, -2, -1, 0, 1, 2, \dots$. At the points $x = \dots, -2, -1, 0, 1, 2, \dots$, $\Phi_\xi(x)$ has jumps equal to the corresponding probabilities $P_\xi(x)$. (b) A typical distribution function of a continuous random variable. Any continuous monotonic function $\Phi_\xi(x)$ such that $\lim_{x \rightarrow -\infty} \Phi_\xi(x) = 0$, $\lim_{x \rightarrow +\infty} \Phi_\xi(x) = 1$ can serve as the distribution function of a continuous random variable ξ .³

variables ξ_1 and ξ_2 . The probability of any event of the type $\{(\xi_1, \xi_2) \in B\}$, i.e., the “probability of the random point $\xi = (\xi_1, \xi_2)$ falling in a given set B ,” is given by

$$P\{(\xi_1, \xi_2) \in B\} = \sum_{(x_1, x_2) \in B} p_{\xi_1, \xi_2}(x_1, x_2),$$

where the summation is over all possible values x_1, x_2 of the random variables ξ_1, ξ_2 such that the point (x_1, x_2) lies in B . Next suppose ξ_1 and ξ_2 are continuous. Then by the *joint probability density* of ξ_1 and ξ_2 , we mean a

³ It should be noted that there are random variables which are neither discrete nor continuous but a “mixture of both.” There are also continuous distribution functions more general than (4.3).

function $p_{\xi_1, \xi_2}(x_1, x_2)$ of two variables x_1 and x_2 such that the probability of any event of the type $\{(\xi_1, \xi_2) \in B\}$ is given by

$$\mathbf{P}\{(\xi_1, \xi_2) \in B\} = \iint_B p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \quad (4.6)$$

(the integral is over B).

Given a family of random variables $\xi_1, \dots, \xi_n$, suppose the events $\{x'_k < \xi_k < x''_k\}$, $k = 1, \dots, n$ are independent for arbitrary x'_k and x''_k ($x'_k < x''_k$). Then the random variables $\xi_1, \dots, \xi_n$ are said to be (*statistically*) *independent*. Given an infinite sequence of random variables $\xi_1, \xi_2, \dots$, suppose the random variables $\xi_1, \dots, \xi_n$ are independent for every n , or equivalently that the events $\{x'_k < \xi_k < x''_k\}$, $k = 1, 2, \dots$ are independent for arbitrary x'_k and x''_k . Then $\xi_1, \xi_2, \dots$ is said to be a *sequence of independent random variables*.

Suppose two random variables ξ_1 and ξ_2 are independent. Then clearly their joint probability distribution (4.5) is such that

$$P_{\xi_1, \xi_2}(x_1, x_2) = P_{\xi_1}(x_1)P_{\xi_2}(x_2) \quad (4.7)$$

if ξ_1 and ξ_2 are discrete, and

$$p_{\xi_1, \xi_2}(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2) \quad (4.7')$$

if ξ_1 and ξ_2 are continuous. In (4.7'), $p_{\xi_1}(x_1)$ is the probability density of ξ_1 and $p_{\xi_2}(x_2)$ that of ξ_2 , while $p_{\xi_1, \xi_2}(x_1, x_2)$ is the joint probability density of ξ_1 and ξ_2 figuring in (4.6).

Example 1 (The uniform distribution). Suppose a point ξ is "tossed at random" onto the interval $[a, b]$. This means that the probability of ξ falling in a subinterval $[x', x''] \subset [a, b]$ does not depend on the location of $[x', x'']$. Hence the probability of ξ falling in $[x', x'']$ is proportional to the length $x'' - x'$.⁴ More exactly, we have

$$\mathbf{P}\{x' < \xi < x''\} = \frac{x'' - x'}{b - a} = \int_{x'}^{x''} \frac{dx}{b - a},$$

since then the probability of ξ falling in $[a, b]$ itself is

$$\mathbf{P}\{a < \xi < b\} = \int_a^b \frac{dx}{b - a} = 1,$$

as it must be. Clearly, ξ is a continuous random variable, with probability

⁴ Let $f(s)$ be the probability of ξ falling in a subinterval of length s . Then clearly $f(s + t) = f(s) + f(t)$. But it can be shown that any function $f(s)$ satisfying this equation is either of the form ks (k a constant) or else unbounded in every interval (see W. Feller, *op. cit.*, p. 459).

density

$$p_{\xi}(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{if } x < a \text{ or } x > b. \end{cases}$$

Such a random variable is said to have a *uniform distribution*.

Example 2. Suppose two points ξ_1 and ξ_2 are tossed at random and *independently* onto a line segment of length L . What is the probability that the distance between the two points does not exceed l ?

Solution. Imagine that ξ_1 falls in an interval $[0, L]$ of the x_1 -axis, while ξ_2 falls in an interval $[0, L]$ of the x_2 -axis, perpendicular to the x_1 -axis as in Figure 5. Then the desired probability is just the probability that a point $\xi = (\xi_1, \xi_2)$ tossed at random onto the square $0 < x_1, x_2 < L$ will fall in the

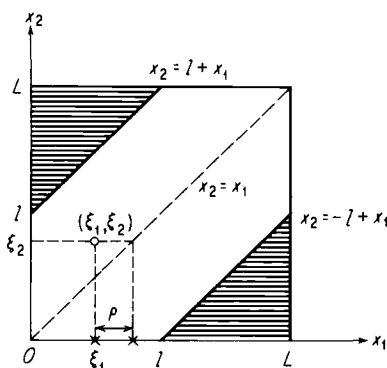


FIGURE 5

region B bounded by the lines $x_2 = l + x_1$ and $x_2 = -l + x_1$ (B is the unshaded region in Figure 5).⁵ By hypothesis, the random variables ξ_1 and ξ_2 are independent and are both uniformly distributed in $[0, L]$, i.e., both have probability density

$$p(x) = \frac{1}{L}, \quad 0 \leq x \leq L.$$

Hence, by (4.6), the joint probability density of the independent random variables ξ_1 and ξ_2 is just

$$p_{\xi_1, \xi_2}(x_1, x_2) = \frac{1}{L^2}, \quad 0 \leq x_1, x_2 \leq L.$$

⁵ Note that $|\xi_1 - \xi_2|$ is the horizontal distance between the point (ξ_1, ξ_2) and the line $x_2 = x_1$. This is the distance ρ shown in Figure 5, from which it is apparent that $\rho \leq l$ if and only if (ξ_1, ξ_2) lies in B .

Therefore the probability of the random point $\xi = (\xi_1, \xi_2)$ falling in the region B is given by

$$P\{(\xi_1, \xi_2) \in B\} = \iint_B \frac{dx_1 dx_2}{L^2} = \frac{2Ll - l^2}{L^2},$$

since $L^2 - 2 \cdot \frac{1}{2}(L-l)^2 = 2Ll - l^2$ is the area of B (the square minus the two shaded triangles).

Example 3 (Buffon's needle problem). Suppose a needle is tossed at random onto a plane ruled with parallel lines a distance L apart, where by a "needle" we mean a line segment of length $l < L$. What is the probability of the needle intersecting one of the parallel lines?

Solution. Let ξ_1 be the angle between the needle and the direction of the rulings, and let ξ_2 be the distance between the bottom point of the needle and the nearest line above this point [see Figure 6(a)]. Then the conditions

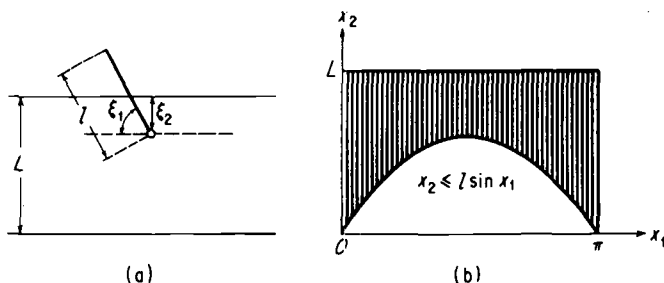


FIGURE 6

of the "needle tossing experiment" are such that the random variable ξ_1 is uniformly distributed in the interval $[0, \pi]$, while the random variable ξ_2 is uniformly distributed in the interval $[0, L]$. Hence, assuming that the random variables ξ_1 and ξ_2 are independent, we find that their joint probability density is

$$p_{\xi_1, \xi_2}(x_1, x_2) = \frac{1}{\pi L}, \quad 0 \leq x_1 < \pi, \quad 0 \leq x_2 \leq L.$$

The event consisting of the needle intersecting one of the rulings occurs if and only if

$$\xi_2 \leq l \sin \xi_1,$$

i.e., if and only if the corresponding point $\xi = (\xi_1, \xi_2)$ falls in the region B , where B is the part of the rectangle $0 \leq x_1 < \pi$, $0 \leq x_2 \leq L$ lying between the x_1 -axis and the curve $x_2 = l \sin x_1$ [B is the unshaded region in Figure

6(b)]. Hence, by the general formula (4.6),

$$\mathbf{P}\{(\xi_1, \xi_2) \in B\} = \iint_B \frac{dx_1 dx_2}{\pi L} = \frac{2l}{\pi L}, \quad (4.8)$$

where

$$l \int_0^\pi \sin x_1 dx_1 = 2l$$

is the area of B .

In deducing (4.8), we have assumed that ξ_1 and ξ_2 are independent random variables. This assumption can be tested experimentally. In fact, according to (4.8), if the needle is repeatedly tossed onto the ruled plane, then the frequency of the event A , consisting of the needle intersecting one of the rulings, must be approximately $2l/\pi L$. Suppose the needle is tossed n times, and let $n(A)$ be the number of times A occurs, so that $n(A)/n$ is the relative frequency of the event A . Then

$$\frac{n(A)}{n} \sim \frac{2l}{\pi L}$$

for large n , as discussed on p. 3. Hence

$$\frac{2l}{L} \frac{n}{n(A)}$$

should be a good approximation to $\pi = 3.14 \dots$ for large n . This actually turns out to be the case.⁶

Example 4. Given two independent random variables ξ_1 and ξ_2 , with probability densities $p_{\xi_1}(x_1)$ and $p_{\xi_2}(x_2)$, find the probability density of the random variable

$$\eta = \xi_1 + \xi_2.$$

Solution. By (4.7'), the joint probability distribution of ξ_1 and ξ_2 equals $p_{\xi_1}(x_1)p_{\xi_2}(x_2)$, and hence, by (4.6),

$$\begin{aligned} P\{y' \leq \eta \leq y''\} &= \iint_{y' \leq x_1 + x_2 \leq y''} p_{\xi_1}(x_1)p_{\xi_2}(x_2) dx_1 dx_2 \\ &= \int_{y'}^{y''} \left[\int_{-\infty}^{\infty} p_{\xi_1}(y-x)p_{\xi_2}(x) dx \right] dy. \end{aligned}$$

Therefore the probability density of the random variable η is given by the expression

$$p_\eta(y) = \int_{-\infty}^{\infty} p_{\xi_1}(y-x)p_{\xi_2}(x) dx,$$

called the *composition* or *convolution* of the functions p_{ξ_1} and p_{ξ_2} .

⁶ See J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill Book Co., Inc., New York (1937), p. 113.

For example, suppose ξ_1 and ξ_2 are both uniformly distributed in the interval $[0, 1]$, so that they both have the probability density

$$p(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1, \\ 0 & \text{if } x < 0 \text{ or } x > 1. \end{cases}$$

Then

$$p_\eta(y) = \begin{cases} \int_0^y dx = y & \text{if } 0 \leq y \leq 1, \\ \int_{y-1}^1 dx = 2 - y & \text{if } 1 \leq y \leq 2, \\ 0 & \text{if } y < 0 \text{ or } y > 2. \end{cases}$$

The graph of the density $p_\eta(y)$ is triangular in shape, as shown in Figure 7.

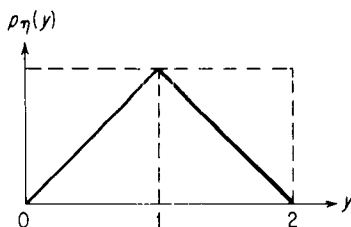


FIGURE 7

8. Mathematical Expectation

By the *mathematical expectation* or *mean value* of a discrete random variable ξ , denoted by $E\xi$, we mean the quantity

$$E\xi = \sum_{-\infty}^{\infty} x P_\xi(x), \quad (4.9)$$

provided that the series converges absolutely.⁷ Here the summation has the same meaning as on p. 37, and, as usual, $P_\xi(x) = P\{\xi = x\}$. Given a discrete random variable ξ , consider the new random variable $\eta = \varphi(\xi)$, where $\varphi(x)$ is some function of x . Then the mathematical expectation of η is given in terms of the probability distribution of ξ by the formula

$$E\eta = E\varphi(\xi) = \sum_{-\infty}^{\infty} \varphi(x) P_\xi(x). \quad (4.10)$$

⁷ I.e., provided that $\sum_{-\infty}^{\infty} |x| P_\xi(x) < \infty$.

In fact, η is a discrete random variable taking only the values $y = \varphi(x)$, where x ranges over all possible values of the random variable ξ . Therefore⁸

$$\mathbf{P}\{\eta = y\} = \sum_{x: \varphi(x)=y} P_{\xi}(x),$$

where the summation is over all x such that $\varphi(x) = y$, and hence

$$\mathbf{E}\eta = \sum_{-\infty}^{\infty} y \mathbf{P}\{\eta = y\} = \sum_{-\infty}^{\infty} \sum_{x: \varphi(x)=y} P_{\xi}(x) = \sum_{-\infty}^{\infty} \varphi(x) P_{\xi}(x),$$

as asserted.

More generally, let $\varphi(\xi_1, \xi_2)$ be a random variable which is a function of two random variables ξ_1 and ξ_2 , with joint probability distribution $P_{\xi_1, \xi_2}(x_1, x_2)$. Then it is easily verified that $\varphi(\xi_1, \xi_2)$ has the mathematical expectation

$$\mathbf{E}\varphi(\xi_1, \xi_2) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \varphi(x_1, x_2) P_{\xi_1, \xi_2}(x_1, x_2). \quad (4.11)$$

It is clear from (4.9) that

$$\text{a) } \mathbf{E}1 = 1;$$

$$\text{b) } \mathbf{E}(c\xi) = c\mathbf{E}\xi \text{ for an arbitrary constant } c;$$

$$\text{c) } |\mathbf{E}\xi| \leq \mathbf{E}|\xi|. \quad (4.12)$$

Moreover, it follows from (4.11) that

$$\text{d) } \mathbf{E}(\xi_1 + \xi_2) = \mathbf{E}\xi_1 + \mathbf{E}\xi_2 \text{ for arbitrary random variables } \xi_1 \text{ and } \xi_2 \\ \text{with mathematical expectations } \mathbf{E}\xi_1 \text{ and } \mathbf{E}\xi_2;$$

$$\text{e) } \xi > 0 \text{ implies } \mathbf{E}\xi \geq 0, \text{ and more generally}$$

$$\xi_1 < \xi_2 \text{ implies } \mathbf{E}\xi_1 < \mathbf{E}\xi_2; \quad (4.13)$$

f) If ξ_1 and ξ_2 are independent random variables, then

$$\mathbf{E}\xi_1\xi_2 = \mathbf{E}\xi_1 \mathbf{E}\xi_2. \quad (4.14)$$

For example, to prove (4.14), we write $\varphi(\xi_1, \xi_2) = \xi_1\xi_2$. Then, for independent ξ_1 and ξ_2 , (4.11) implies

$$\begin{aligned} \mathbf{E}(\xi_1\xi_2) &= \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} x_1x_2 P_{\xi_1}(x_1) P_{\xi_2}(x_2) \\ &= \sum_{-\infty}^{\infty} x_1 P_{\xi_1}(x_1) \sum_{-\infty}^{\infty} x_2 P_{\xi_2}(x_2) = \mathbf{E}\xi_1 \mathbf{E}\xi_2. \end{aligned}$$

To define the mathematical expectation of a *continuous* random variable ξ , we first approximate ξ by a sequence of discrete random variables ξ_n ,

⁸ The colon should be read as "such that."

$n = 1, 2, \dots$. Let ε_n , $n = 1, 2, \dots$ be a sequence of positive numbers converging to zero. Then for each $n = 1, 2, \dots$, let

$$\dots, x_{-2,n}, x_{-1,n}, x_{0,n}, x_{1,n}, x_{2,n}, \dots \quad (4.15)$$

be an infinite set of distinct points such that⁹

$$\sup_k |x_{k,n} - x_{k-1,n}| = \varepsilon_n, \quad (4.16)$$

and let ξ_n be a discrete random variable such that

$$\xi_n = x_{k,n} \quad \text{if} \quad x_{k-1,n} < \xi \leq x_{k,n}.$$

It follows that

$$|\xi - \xi_n| \leq \varepsilon_n,$$

and hence

$$|\xi_m - \xi_n| \leq |\xi_m - \xi_n| + |\xi_n - \xi| \leq \varepsilon_m + \varepsilon_n \rightarrow 0$$

as $m, n \rightarrow \infty$. Therefore, by (4.12) and (4.13),

$$|E\xi_m - E\xi_n| = |E(\xi_m - \xi_n)| \leq E|\xi_m - \xi_n| \leq \varepsilon_m + \varepsilon_n \rightarrow 0$$

as $m, n \rightarrow \infty$ (provided $E\xi_n$ exists for all n). But then

$$\lim_{n \rightarrow \infty} E\xi_n$$

exists, by the Cauchy convergence criterion. This limit is called the *mathematical expectation* or *mean value* of the continuous random variable ξ , again denoted by $E\xi$. Clearly,

$$E\xi = \lim_{n \rightarrow \infty} \sum_{-\infty}^{\infty} x_{k,n} P\{x_{k-1,n} < \xi \leq x_{k,n}\}.$$

Suppose ξ has the probability density $p_\xi(x)$. Then, choosing the points (4.15) to be continuity points of $p_\xi(x)$, we have

$$\begin{aligned} \sum_{-\infty}^{\infty} x_{k,n} P\{x_{k-1,n} < \xi \leq x_{k,n}\} &= \sum_{-\infty}^{\infty} x_{k,n} \int_{x_{k-1,n}}^{x_{k,n}} p_\xi(x) dx \\ &\sim \sum_{-\infty}^{\infty} x_{k,n} p_\xi(x_{k,n}) (x_{k,n} - x_{k-1,n}), \end{aligned}$$

and hence

$$E\xi = \int_{-\infty}^{\infty} x p_\xi(x) dx \quad (4.17)$$

[compare this with (4.9)]. For continuous random variables of the form

⁹ The symbol $\sup$ denotes the supremum or least upper bound. Therefore the left-hand side of (4.16) is the least upper bound of all the differences $|x_{k,n} - x_{k-1,n}|$, $k = \dots, -2, -1, 0, 1, 2, \dots$. Thus (4.16) means that no two of the points (4.15) are more than ε_n apart. Note also that any closed interval of length ε_n contains at least two of the points (4.15).

$\eta = \varphi(\xi)$ and $\eta = \varphi(\xi_1, \xi_2)$, we have

$$E\varphi(\xi) = \int_{-\infty}^{\infty} \varphi(x)p_{\xi}(x) dx \quad (4.18)$$

and

$$E\varphi(\xi_1, \xi_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x_1, x_2)p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2, \quad (4.19)$$

by analogy with (4.10) and (4.11), where $p_{\xi_1, \xi_2}(x_1, x_2)$ is the joint probability density of the random variables ξ_1 and ξ_2 . It is easily verified that properties a)–f) of the mathematical expectation continue to hold for continuous random variables (the details are left as an exercise).

Remark. Other synonyms for the mathematical expectation of a random variable ξ , discrete or continuous, are the *expected value* of ξ or the *average (value)* of ξ . The mathematical expectation and mean value are often simply called the “expectation” and “mean,” respectively.

Example Let ξ be a random variable uniformly distributed in the interval $[a, b]$, i.e., let ξ have the probability density

$$p_{\xi}(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b, \\ 0 & \text{if } x < a \text{ or } x > b. \end{cases}$$

Then the mathematical expectation of ξ is

$$E\xi = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}.$$

A random variable of the form $\eta = \xi_1 + i\xi_2$ involving two real random variables ξ_1 and ξ_2 (the real and imaginary parts of η) is called a *complex random variable*. The mathematical expectation of $\eta = \xi_1 + i\xi_2$ is defined as

$$E\eta = E\xi_1 + iE\xi_2.$$

It is easy to see that formulas (4.10) and (4.18) remain valid for the case where $\varphi(\xi)$ is a complex-valued function of a real random variable ξ , and that (4.11) and (4.19) remain valid for the case where $\varphi(\xi_1, \xi_2)$ is a complex-valued function of two real random variables ξ_1 and ξ_2 . In particular, let $\varphi_1(\xi_1)$ and $\varphi_2(\xi_2)$ be complex-valued functions of two *independent* real random variables ξ_1 and ξ_2 . Then, choosing $\varphi(\xi_1, \xi_2) = \varphi_1(\xi_1)\varphi_2(\xi_2)$ in (4.11) or (4.19), we deduce the formula

$$E[\varphi_1(\xi_1)\varphi_2(\xi_2)] = E\varphi_1(\xi_1)E\varphi_2(\xi_2), \quad (4.20)$$

which generalizes (4.14).

9. Chebyshev's Inequality. The Variance and Correlation Coefficient

By the *mean square value* of a (real) random variable ξ is meant the quantity $E\xi^2$, equal to

$$E\xi^2 = \sum_{-\infty}^{\infty} x^2 P_{\xi}(x)$$

if ξ is discrete, or

$$E\xi^2 = \int_{-\infty}^{\infty} x^2 p_{\xi}(x) dx$$

if ξ is continuous.¹⁰ Given any random variable ξ and any number $\epsilon > 0$, let

$$\xi_1 = \begin{cases} 0 & \text{if } |\xi| < \epsilon, \\ \epsilon^2 & \text{if } |\xi| > \epsilon. \end{cases}$$

Then obviously $\xi_1 \leq \xi^2$, and hence, by (4.13),

$$E\xi_1 \leq E\xi^2,$$

or equivalently

$$\epsilon^2 P\{|\xi| > \epsilon\} \leq E\xi^2,$$

since clearly

$$E\xi_1 = \epsilon^2 P\{|\xi| > \epsilon\}.$$

It follows that

$$P\{|\xi| > \epsilon\} \leq \frac{1}{\epsilon^2} E\xi^2, \quad (4.21)$$

a result known as *Chebyshev's inequality*. According to (4.21), if $E\xi^2/\epsilon^2 < \delta$, then $P\{|\xi| > \epsilon\} < \delta$, and hence $P\{|\xi| \leq \epsilon\} > 1 - \delta$. Therefore, if δ is small, it is highly probable that $|\xi| \leq \epsilon$. In particular, if $E\xi^2 = 0$, then $P\{|\xi| > \epsilon\} = 0$ for every $\epsilon > 0$, and hence $\xi = 0$ with probability 1.

By the *variance* or *dispersion* of a random variable ξ , denoted by $D\xi$, we mean the mean square value $E(\xi - a)^2$ of the difference $\xi - a$, where $a = E\xi$ is the mean value of ξ . It follows from

$$E(\xi - a)^2 = E\xi^2 - 2aE\xi + a^2 = E\xi^2 - 2a^2 + a^2$$

that

$$D\xi = E\xi^2 - a^2.$$

Obviously

$$D1 = 0,$$

and

$$D(c\xi) = c^2 D\xi$$

for an arbitrary constant c .

¹⁰ It is assumed, of course, that $E\xi^2$ exists. This is not always the case (see e.g., Problem 24, p. 53).

If ξ_1 and ξ_2 are independent random variables, then

$$D(\xi_1 + \xi_2) = D\xi_1 + D\xi_2.$$

In fact, if $a_1 = E\xi_1$ and $a_2 = E\xi_2$, then, by (4.14),

$$E(\xi_1 - a_1)(\xi_2 - a_2) = E(\xi_1 - a_1)E(\xi_2 - a_2) = 0, \quad (4.22)$$

and hence

$$\begin{aligned} D(\xi_1 + \xi_2) &= E(\xi_1 + \xi_2 - a_1 - a_2)^2 \\ &= E(\xi_1 - a_1)^2 + 2E(\xi_1 - a_1)(\xi_2 - a_2) + E(\xi_2 - a_2)^2 \\ &= E(\xi_1 - a_1)^2 + E(\xi_2 - a_2)^2 = D\xi_1 + D\xi_2. \end{aligned}$$

Given two random variables ξ_1 and ξ_2 , we now consider the problem of finding the linear expression of the form $c_1 + c_2\xi_2$, involving constants c_1 and c_2 , such that $c_1 + c_2\xi_2$ is the best “mean square approximation” to ξ_1 , in the sense that

$$E(\xi_1 - c_1 - c_2\xi_2)^2 = \min_{c_1, c_2} E(\xi_1 - c_1 - c_2\xi_2)^2, \quad (4.23)$$

where the minimum is taken with respect to all c_1 and c_2 . To solve this problem, we let

$$\begin{aligned} a_1 &= E\xi_1, & \sigma_1^2 &= D\xi_1, \\ a_2 &= E\xi_2, & \sigma_2^2 &= D\xi_2, \end{aligned} \quad (4.24)$$

and introduce the quantity

$$r = \frac{E(\xi_1 - a_1)(\xi_2 - a_2)}{\sigma_1\sigma_2}, \quad (4.25)$$

called the *correlation coefficient* of the random variables ξ_1 and ξ_2 . Going over for convenience to the “normalized” random variables

$$\eta_1 = \frac{\xi_1 - a_1}{\sigma_1}, \quad \eta_2 = \frac{\xi_2 - a_2}{\sigma_2},$$

we find that

$$\min_{c_1, c_2} E(\xi_1 - c_1 - c_2\xi_2)^2 = \sigma_1^2 \min_{c_1, c_2} E(\eta_1 - c_1 - c_2\eta_2)^2 \quad (4.26)$$

(why?). Clearly,

$$E\eta_1 = E\eta_2 = 0, \quad D\eta_1 = E\eta_1^2 = D\eta_2 = E\eta_2^2 = 1, \quad E\eta_1\eta_2 = r,$$

$$E(\eta_1 - r\eta_2)\eta_2 = E\eta_1\eta_2 - rE\eta_2^2 = 0,$$

$$E(\eta_1 - r\eta_2)^2 = E\eta_1^2 - 2rE\eta_1\eta_2 + r^2E\eta_2^2 = 1 - r^2,$$

and hence

$$\begin{aligned} E(\eta_1 - c_1 - c_2\eta_2)^2 &= E[(\eta_1 - r\eta_2) - c_1 + (r - c_2)\eta_2]^2 \\ &= (1 - r^2) + c_1^2 + (r - c_2)^2 \end{aligned}$$

for arbitrary c_1 and c_2 . It follows that the minimum of $E(\eta_1 - c_1 - c_2\eta_2)^2$ is achieved for $c_1 = 0$, $c_2 = r$, and is equal to

$$\min_{c_1, c_2} E(\eta_1 - c_1 - c_2\eta_2)^2 = 1 - r^2. \quad (4.27)$$

But

$$\eta_1 - r\eta_2 = \frac{1}{\sigma_1} \left[\xi_1 - a_1 - r \frac{\sigma_1}{\sigma_2} (\xi_2 - a_2) \right]$$

in terms of the original random variables ξ_1 and ξ_2 . Therefore

$$\hat{c}_1 + \hat{c}_2 \xi_2 = a_1 + r \frac{\sigma_1}{\sigma_2} (\xi_2 - a_2)$$

is the minimizing linear expression figuring in (4.23), where a_1 , a_2 , σ_1^2 , σ_2^2 and r are defined by (4.24) and (4.25).

If ξ_1 and ξ_2 are *independent*, then, by (4.22),

$$r = \frac{E(\xi_1 - a_1)(\xi_2 - a_2)}{\sigma_1 \sigma_2} = \frac{E(\xi_1 - a_1)E(\xi_2 - a_2)}{\sigma_1 \sigma_2} = 0.$$

It is clear from (4.27) that r lies in the interval $-1 < r < 1$. Moreover, if $r = \pm 1$, then the random variable ξ_1 is simply a linear expression of the form

$$\xi_1 = \hat{c}_1 + \hat{c}_2 \xi_2.$$

In fact, if $r = \pm 1$, then, by (4.26) and (4.27), the mean square value of $\xi_1 - \hat{c}_1 - \hat{c}_2 \xi_2$ is just

$$E(\xi_1 - \hat{c}_1 - \hat{c}_2 \xi_2)^2 = \sigma_1^2(1 - r^2) = 0,$$

and hence $\xi_1 - \hat{c}_1 - \hat{c}_2 \xi_2 = 0$ with probability 1 (why?).

The above considerations seem to suggest the use of r as a measure of the extent to which the random variables ξ_1 and ξ_2 are dependent. However, although suitable in some situations (see Problem 15, p. 67), this use of r is not justified in general (see Problem 19, p. 53).¹¹

PROBLEMS

1. A motorist encounters four consecutive traffic lights, each equally likely to be red or green. Let ξ be the number of green lights passed by the motorist before being stopped by a red light. What is the probability distribution of ξ ?
2. Give an example of two distinct random variables with the same distribution function.
3. Find the distribution function of the uniformly distributed random variable ξ considered in Example 1, p. 40.

¹¹ See also W. Feller, *op. cit.*, p. 236.

4. A random variable ξ has probability density

$$p_{\xi}(x) = \frac{a}{x^2 + 1} \quad (-\infty < x < \infty).$$

Find

- a) The constant a ; b) The distribution function of ξ ;
c) The probability $P\{-1 < \xi < 1\}$.

Ans. a) $\frac{1}{\pi}$; b) $\frac{1}{2} + \frac{1}{\pi} \arctan x$; c) $\frac{1}{2}$.

5. A random variable ξ has probability density

$$p_{\xi}(x) = \begin{cases} ax^2 e^{-kx} & \text{if } 0 < x < \infty, \\ 0 & \text{otherwise} \end{cases}$$

($k > 0$). Find

- a) The constant a ; b) The distribution function of ξ ;
c) The probability $P\{0 < \xi < 1/k\}$.

6. A random variable ξ has distribution function

$$\Phi_{\xi}(x) = a + b \arctan \frac{x}{2} \quad (-\infty < x < \infty).$$

Find

- a) The constants a and b ; b) The probability density of ξ .

7. Two nonoverlapping circular disks of radius r are painted on a circular table of radius R . A point is then "tossed at random" onto the table. What is the probability of the point falling in one of the disks?

Ans. $2(r/R)^2$.

8. What is the probability that two randomly chosen numbers between 0 and 1 will have a sum no greater than 1 and a product no greater than $\frac{2}{9}$?

Ans. $\frac{1}{3} + \frac{2}{9} \int_{1/3}^{2/3} \frac{dx}{x} = \frac{1}{3} + \frac{2}{9} \ln 2 \approx 0.49$.

9. Given two independent random variables ξ_1 and ξ_2 , with probability densities

$$p_{\xi_1}(x) = \begin{cases} \frac{1}{2} e^{-x/2} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases} \quad p_{\xi_2}(x) = \begin{cases} \frac{1}{3} e^{-x/3} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$$

find the probability density of the random variable $\eta = \xi_1 + \xi_2$.

Ans. $p_{\eta}(x) = \begin{cases} e^{-x/3}(1 - e^{-x/6}) & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$

10. Given three independent random variables ξ_1 , ξ_2 and ξ_3 , each uniformly distributed in the interval $[0, 1]$, find the probability density of the random variable $\xi_1 + \xi_2 + \xi_3$.

Hint. The probability density of $\xi_1 + \xi_2$ (say) was found in Example 4, p. 43.

11. A random variable ξ takes the values $1, 2, \dots, n, \dots$ with probabilities

$$\frac{1}{3}, \frac{1}{3^2}, \dots, \frac{1}{3^n}, \dots$$

Find $E\xi$.

12. Balls are drawn from an urn containing w white balls and b black balls until a white ball appears. Find the mean value m and variance σ^2 of the number of black balls drawn, assuming that each ball is replaced after being drawn.

$$\text{Ans. } m = \frac{b}{w}, \sigma^2 = \frac{b(w+b)}{w^2}.$$

13. Find the mean and variance of the random variable ξ with probability density

$$p_\xi(x) = \frac{1}{2}e^{-|x|} \quad (-\infty < x < \infty).$$

$$\text{Ans. } E\xi = 0, D\xi = 2.$$

14. Find the mean and variance of the random variable ξ with probability density

$$p_\xi(x) = \begin{cases} \frac{1}{2b} & \text{if } |x - a| \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{Ans. } E\xi = a, D\xi = b^2/3.$$

15. The distribution function of a random variable ξ is

$$\Phi_\xi(x) = \begin{cases} 0 & \text{if } x \leq -1, \\ a + b \arcsin x & \text{if } -1 \leq x \leq 1, \\ 1 & \text{if } x \geq 1. \end{cases}$$

Find $E\xi$ and $D\xi$.

Hint. First determine a and b .

$$\text{Ans. } E\xi = 0, D\xi = \frac{1}{2}.$$

16. Let ξ be the number of spots obtained in throwing an unbiased die. Find the mean and variance of ξ .

$$\text{Ans. } E\xi = \frac{7}{2}, D\xi = \frac{35}{12}.$$

17. In the preceding problem, what is the probability P of ξ deviating from $E\xi$ by more than $\frac{5}{2}$? Show that Chebyshev's inequality gives only a very crude estimate of P .

18. Prove that if ξ is a random variable such that $Ee^{a\xi}$ exists, where a is a positive constant, then

$$P\{\xi > \varepsilon\} < \frac{Ee^{a\xi}}{e^{a\varepsilon}}$$

Hint. Apply Chebyshev's inequality to the random variable $\eta = e^{a\xi/2}$.

19. Let ξ be a random variable taking each of the values $-2, -1, 1$ and 2 with probability $\frac{1}{4}$, and let $\eta = \xi^2$. Prove that ξ and η (although obviously dependent) have correlation coefficient 0.

20. Find the means and variances of two random variables ξ_1 and ξ_2 with joint probability density

$$p_{\xi_1, \xi_2}(x_1, x_2) = \begin{cases} \sin x_1 \sin x_2 & \text{if } 0 < x_1 < \frac{\pi}{2}, \quad 0 < x_2 < \frac{\pi}{2} \\ 0 & \text{otherwise.} \end{cases}$$

What is the correlation coefficient of ξ_1 and ξ_2 ?

21. Find the correlation coefficient r of two random variables ξ_1 and ξ_2 with joint probability density

$$p_{\xi_1, \xi_2}(x_1, x_2) = \begin{cases} \frac{1}{2} \sin(x_1 + x_2) & \text{if } 0 \leq x_1 \leq \frac{\pi}{2}, \quad 0 \leq x_2 \leq \frac{\pi}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{Ans. } r = \frac{\frac{\pi}{2} - 1 - \frac{\pi^2}{16}}{\frac{\pi}{2} - 2 + \frac{\pi^2}{16}} \approx -\frac{1}{4}.$$

22. Given a random variable ξ , let $\varphi(t)$ be a nondecreasing positive function such that $E\varphi(\xi)$ exists. Prove that

$$P\{\xi > t\} < \frac{m}{\varphi(t)}. \quad (4.28)$$

23. Deduce Chebyshev's inequality as a special case of (4.28).

24. Let ξ be a random variable with probability density

$$p_{\xi}(x) = \frac{1}{\pi(1+x^2)} \quad (-\infty < x < \infty).$$

Show that $E\xi$ and $D\xi$ fail to exist.