Mathematical operations over random variables.

Let two random variables X and Y be given:

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 & y_2 & \dots & y_m \\ p'_1 & p'_2 & \dots & p'_m \end{pmatrix}.$$

The product kX of a random variable X on a constant k is the random variable which takes on values kx_i with the same probabilities p_i $(i = 1, 2, .., n)$.

The m-th degree of a random variable X, i.e. X^m is the random variable which takes on values x_i^m with the same probabilities p_i $(i = 1, 2, .., n)$.

Example. Let a random variable X be given: $X = \begin{pmatrix} -2 & 1 & 2 \\ 0{,}5 & 0{,}3 & 0{,}2 \end{pmatrix}$. Find the law of distribution of the random variables: a) $Y = 3X$; b) $Z = X^2$.

Solution: a) The values of the random variable Y will be: $3 \cdot (-2) = -6$; $3 \cdot 1 = 3$; $3 \cdot 2 = 6$ with the same probabilities 0,5; 0,3; 0,2, i.e. $Y = \begin{pmatrix} -6 & 3 & 6 \\ 0{,}5 & 0{,}3 & 0{,}2 \end{pmatrix}$.

b) The values of the random variable Z will be: $(-2)^2 = 4$, $1^2 = 1$, $2^2 = 4$ with the same probabilities 0,5; 0,3; 0,2. Since the value $Z = 4$ can be obtained by squaring the values (-2) with probability 0,5 and $(+2)$ with probability 0,2, under the theorem of addition: $P(Z = 4) = 0{,}5 + 0{,}2 = 0{,}7$. Thus, we have the following law of the random variable Z:

$$Z = \begin{pmatrix} 1 & 4 \\ 0{,}3 & 0{,}7 \end{pmatrix}$$

The sum (*the difference* or *the product*) of random variables X and Y is the random variable which takes on all possible values of kind $x_i + y_j$ ($x_i - y_j$ or $x_i \cdot y_j$) where $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$ with the probabilities p_{ij} that the random variable X will take on the value x_i, and Y – the value y_j:

$$p_{ij} = P[(X = x_i)(Y = y_j)].$$

If random variables X and Y are independent, i.e. any events $X = x_i$, $Y = y_j$ are independent, then by theorem of multiplication of probabilities for independent events

$$p_{ij} = P(X = x_i) \cdot P(Y = y_j) = p_i \cdot p'_j.$$

The distribution of a random variable X contains all of the probabilistic information about X. The entire distribution of X, however, is usually too cumbersome for presenting this information. Summaries of the distribution, such as the average value, or expected value, can be useful for giving people an idea of where we expect X to be without trying to describe the entire distribution. The expected value also plays an important role in the approximation methods

Example 4.1.1

Fair Price for a Stock. An investor is considering whether or not to invest \$18 per share in a stock for one year. The value of the stock after one year, in dollars, will be $18 + X$, where X is the amount by which the price changes over the year. At present X is unknown, and the investor would like to compute an "average value" for X in order to compare the return she expects from the investment to what she would get by putting the \$18 in the bank at 5% interest. ◀

The idea of finding an average value as in Example 4.1.1 arises in many applications that involve a random variable. One popular choice is what we call the *mean* or *expected value* or *expectation*.

The intuitive idea of the mean of a random variable is that it is the weighted average of the possible values of the random variable with the weights equal to the probabilities.

Example 4.1.2

Stock Price Change. Suppose that the change in price of the stock in Example 4.1.1 is a random variable X that can assume only the four different values -2, 0, 1, and 4, and that $\Pr(X = -2) = 0.1$, $\Pr(X = 0) = 0.4$, $\Pr(X = 1) = 0.3$, and $\Pr(X = 4) = 0.2$. Then the weighted avarage of these values is

$$-2(0.1) + 0(0.4) + 1(0.3) + 4(0.2) = 0.9.$$

The investor now compares this with the interest that would be earned on \$18 at 5% for one year, which is $18 \times 0.05 = 0.9$ dollars. From this point of view, the price of \$18 seems fair. ◄

The calculation in Example 4.1.2 generalizes easily to every random variable that assumes only finitely many values. Possible problems arise with random variables that assume more than finitely many values, especially when the collection of possible values is unbounded.

Definition 4.1.1 **Mean of Bounded Discrete Random Variable.** Let X be a bounded discrete random variable whose p.f. is f. The *expectation of* X, denoted by $E(X)$, is a number defined as follows:

$$E(X) = \sum_{\text{All } x} x f(x). \tag{4.1.1}$$

The expectation of X is also referred to as the *mean of* X or the *expected value of* X.

In Example 4.1.2, $E(X) = 0.9$. Notice that 0.9 is not one of the possible values of X in that example. This is typically the case with discrete random variables.

Example 4.1.3 **Bernoulli Random Variable.** Let X have the Bernoulli distribution with parameter p, that is, assume that X takes only the two values 0 and 1 with $\Pr(X = 1) = p$. Then the mean of X is

$$E(X) = 0 \times (1 - p) + 1 \times p = p. \qquad \blacktriangleleft$$

If X is unbounded, it might still be possible to define $E(X)$ as the weighted average of its possible values. However, some care is needed.

Definition 4.1.2 **Mean of General Discrete Random Variable.** Let X be a discrete random variable whose p.f. is f. Suppose that at least one of the following sums is finite:

$$\sum_{\text{Positive } x} xf(x), \qquad \sum_{\text{Negative } x} xf(x). \tag{4.1.2}$$

Then the *mean*, *expectation*, or *expected value* of X is said to *exist* and is defined to be

$$E(X) = \sum_{\text{All } x} xf(x). \tag{4.1.3}$$

If both of the sums in (4.1.2) are infinite, then $E(X)$ *does not exist*.

The reason that the expectation fails to exist if both of the sums in (4.1.2) are infinite is that, in such cases, the sum in (4.1.3) is not well-defined. It is known from calculus that the sum of an infinite series whose positive and negative terms both add to infinity either fails to converge or can be made to converge to many different values by rearranging the terms in different orders. We don't want the meaning of expected value to depend on arbitrary choices about what order to add numbers. If only one of two sums in (4.1.3) is infiinte, then the expected value is also infinite with the same sign as that of the sum that is infinite. If both sums are finite, then the sum in (4.1.3) converges and doesn't depend on the order in which the terms are added.

Example 4.1.4 **The Mean of X Does Not Exist.** Let X be a random variable whose p.f. is

$$f(x) = \begin{cases} \dfrac{1}{2|x|(|x|+1)} & \text{if } x = \pm 1, \pm 2, \pm 3, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

It can be verified that this function satisfies the conditions required to be a p.f. The two sums in (4.1.2) are

$$\sum_{x=-1}^{-\infty} x \frac{1}{2|x|(|x|+1)} = -\infty \quad \text{and} \quad \sum_{x=1}^{\infty} x \frac{1}{2x(x+1)} = \infty;$$

hence, $E(X)$ does not exist. ◀

Example 4.1.5 **An Infinite Mean.** Let X be a random variable whose p.f. is

$$f(x) = \begin{cases} \dfrac{1}{x(x+1)} & \text{if } x = 1, 2, 3, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

The sum over negative values in Eq. (4.1.2) is 0, so the mean of X exists and is

$$E(X) = \sum_{x=1}^{\infty} x \frac{1}{x(x+1)} = \infty.$$

We say that the mean of X is *infinite* in this case. ◀

Note: The Expectation of X Depends Only on the Distribution of X. Although $E(X)$ is called the expectation of X, it depends only on the distribution of X. Every two random variables that have the same distribution will have the same expectation even if they have nothing to do with each other. For this reason, we shall often refer to the expectation of a distribution even if we do not have in mind a random variable with that distribution.

Expectations for a Continuous Distribution

The idea of computing a weighted average of the possible values can be generalized to continuous random variables by using integrals instead of sums. The distinction between bounded and unbounded random variables arises in this case for the same reasons.

Definition 4.1.3 **Mean of Bounded Continuous Random Variable.** Let X be a bounded continuous random variable whose p.d.f. is f. The *expectation of* X, denoted $E(X)$, is defined as follows:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx. \tag{4.1.4}$$

Once again, the expectation is also called the *mean* or the *expected value*.

Example 4.1.6

Expected Failure Time. An appliance has a maximum lifetime of one year. The time X until it fails is a random variable with a continuous distribution having p.d.f.

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Example 4.1.6 **Expected Failure Time.** An appliance has a maximum lifetime of one year. The time X until it fails is a random variable with a continuous distribution having p.d.f.

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E(X) = \int_0^1 x(2x)\,dx = \int_0^1 2x^2\,dx = \frac{2}{3}.$$

We can also say that the expectation of the distribution with p.d.f. f is 2/3. ◀

Definition 4.1.4 **Mean of General Continuous Random Variable.** Let X be a continuous random variable whose p.d.f. is f. Suppose that at least one of the following integrals is finite:

$$\int_0^\infty xf(x)dx, \quad \int_{-\infty}^0 xf(x)dx. \tag{4.1.5}$$

Then the *mean*, *expectation*, or *expected value* of X is said to *exist* and is defined to be

$$E(X) = \int_{-\infty}^\infty xf(x)dx. \tag{4.1.6}$$

If both of the integrals in (4.1.5) are infinite, then $E(X)$ *does not exist*.

Example 4.1.7 **Failure after Warranty.** A product has a warranty of one year. Let X be the time at which the product fails. Suppose that X has a continuous distribution with the p.d.f.

$$f(x) = \begin{cases} 0 & \text{for } x < 1, \\ \frac{2}{x^3} & \text{for } x \geq 1. \end{cases}$$

The expected time to failure is then

$$E(X) = \int_1^\infty x \frac{2}{x^3} dx = \int_1^\infty \frac{2}{x^2} dx = 2.$$

◀

Example 4.1.8 **A Mean That Does Not Exist.** Suppose that a random variable X has a continuous distribution for which the p.d.f. is as follows:

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{for } -\infty < x < \infty. \tag{4.1.7}$$

This distribution is called the *Cauchy distribution*. We can verify the fact that $\int_{-\infty}^{\infty} f(x)\, dx = 1$ by using the following standard result from elementary calculus:

$$\frac{d}{dx} \tan^{-1} x = \frac{1}{1+x^2} \quad \text{for } -\infty < x < \infty.$$

The two integrals in (4.1.5) are

$$\int_0^\infty \frac{x}{\pi(1+x^2)} dx = \infty \quad \text{and} \quad \int_{-\infty}^0 \frac{x}{\pi(1+x^2)} dx = -\infty;$$

hence, the mean of X does not exist. ◀

Interpretation of the Expectation

Relation of the Mean to the Center of Gravity The expectation of a random variable or, equivalently, the mean of its distribution can be regarded as being the center of gravity of that distribution. To illustrate this concept, consider, for example, the p.f. sketched in Fig. 4.1. The x-axis may be regarded as a long weightless rod to which weights are attached. If a weight equal to $f(x_j)$ is attached to this rod at each point x_j, then the rod will be balanced if it is supported at the point $E(X)$.

Now consider the p.d.f. sketched in Fig. 4.2. In this case, the x-axis may be regarded as a long rod over which the mass varies continuously. If the density of the rod at each point x is equal to $f(x)$, then the center of gravity of the rod will be located at the point $E(X)$, and the rod will be balanced if it is supported at that point.

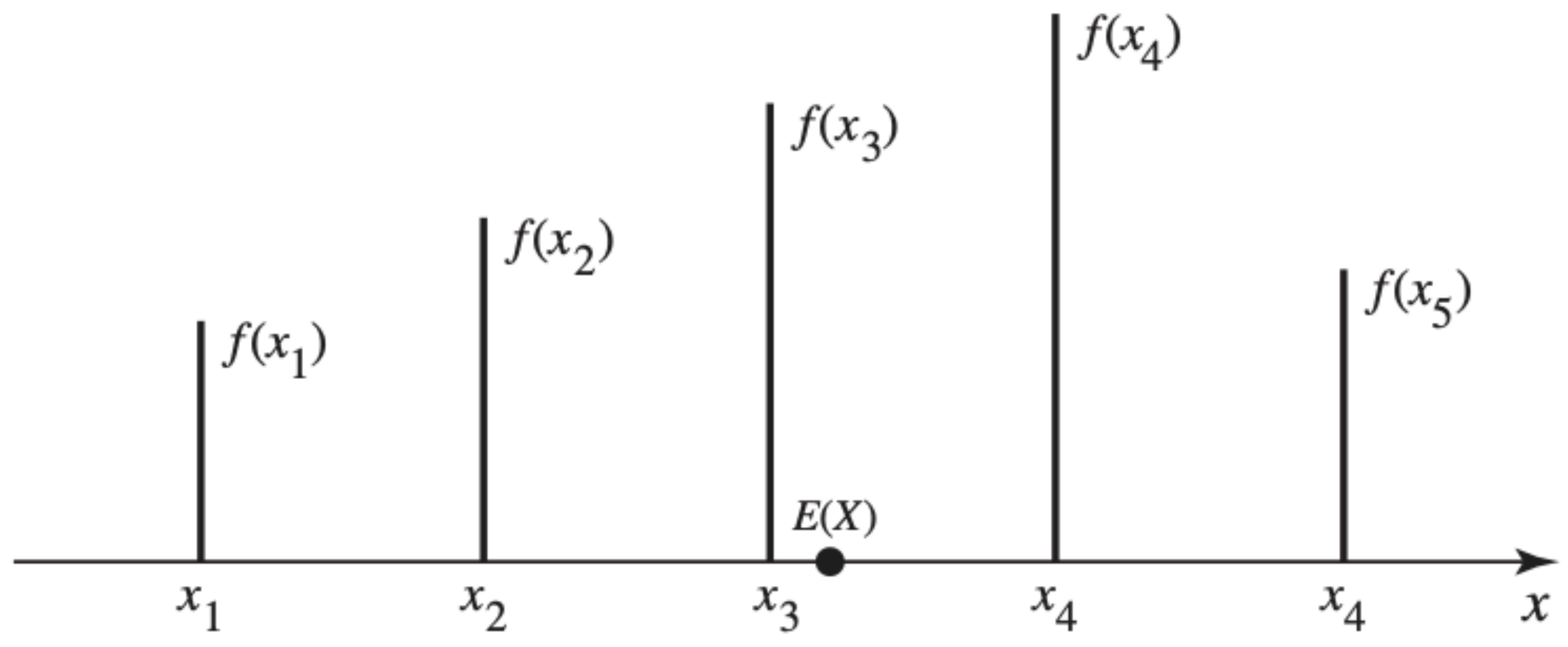


Figure 4.1 The mean of a discrete distribution.

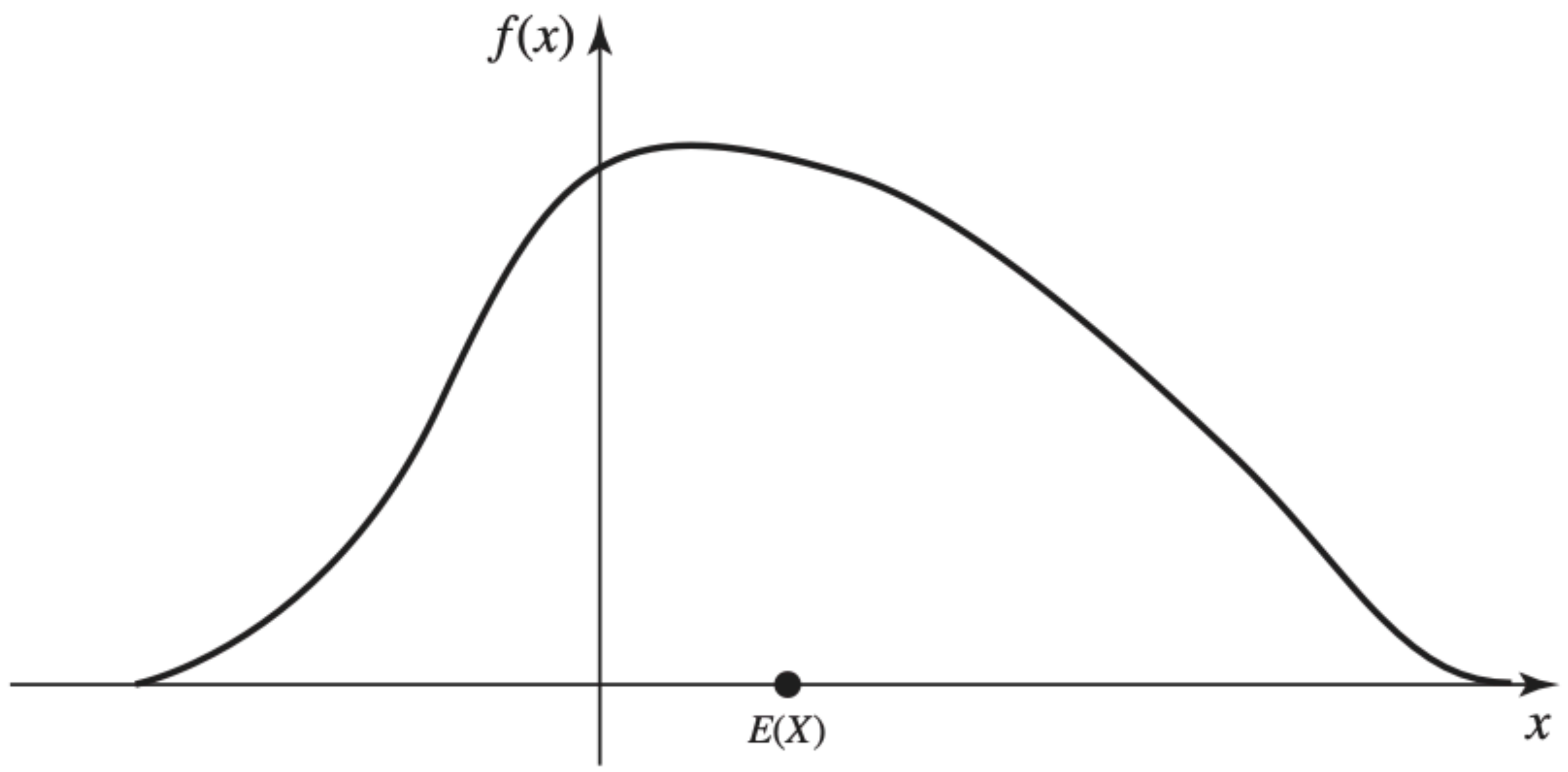


Figure 4.2 The mean of a continuous distribution.

It can be seen from this discussion that the mean of a distribution can be affected greatly by even a very small change in the amount of probability that is assigned to a large value of x. For example, the mean of the distribution represented by the p.f. in Fig. 4.1 can be moved to any specified point on the x-axis, no matter how far from the origin that point may be, by removing an arbitrarily small but positive amount of probability from one of the points x_j and adding this amount of probability at a point far enough from the origin.

Suppose now that the p.f. or p.d.f. f of some distribution is symmetric with respect to a given point x_0 on the x-axis. In other words, suppose that $f(x_0 + \delta) = f(x_0 - \delta)$ for all values of δ. Also assume that the mean E(X) of this distribution exists. In accordance with the interpretation that the mean is at the center of gravity, it follows that E(X) must be equal to x_0, which is the point of symmetry. The following example emphasizes the fact that it is necessary to make certain that the mean E(X) exists before it can be concluded that $E(X) = x_0$.

Example 4.1.9 **The Cauchy Distribution.** Consider again the p.d.f. specified by Eq. (4.1.7), which is sketched in Fig. 4.3. This p.d.f. is symmetric with respect to the point $x = 0$. Therefore, if the mean of the Cauchy distribution existed, its value would have to be 0. However, we saw in Example 4.1.8 that the mean of X does not exist.

The reason for the nonexistence of the mean of the Cauchy distribution is as follows: When the curve $y = f(x)$ is sketched as in Fig. 4.3, its tails approach the x-axis rapidly enough to permit the total area under the curve to be equal to 1. On the other hand, if each value of $f(x)$ is multiplied by x and the curve $y = xf(x)$ is sketched, as in Fig. 4.4, the tails of this curve approach the x-axis so slowly that the total area between the x-axis and each part of the curve is infinite. ◀

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{for } -\infty < x < \infty. \tag{4.1.7}$$

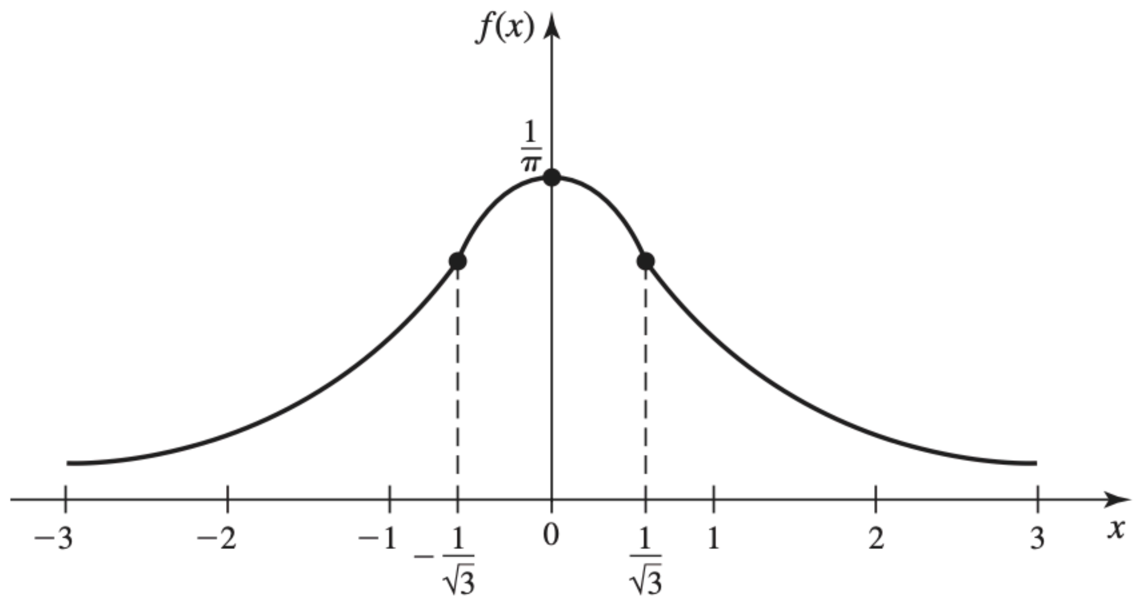
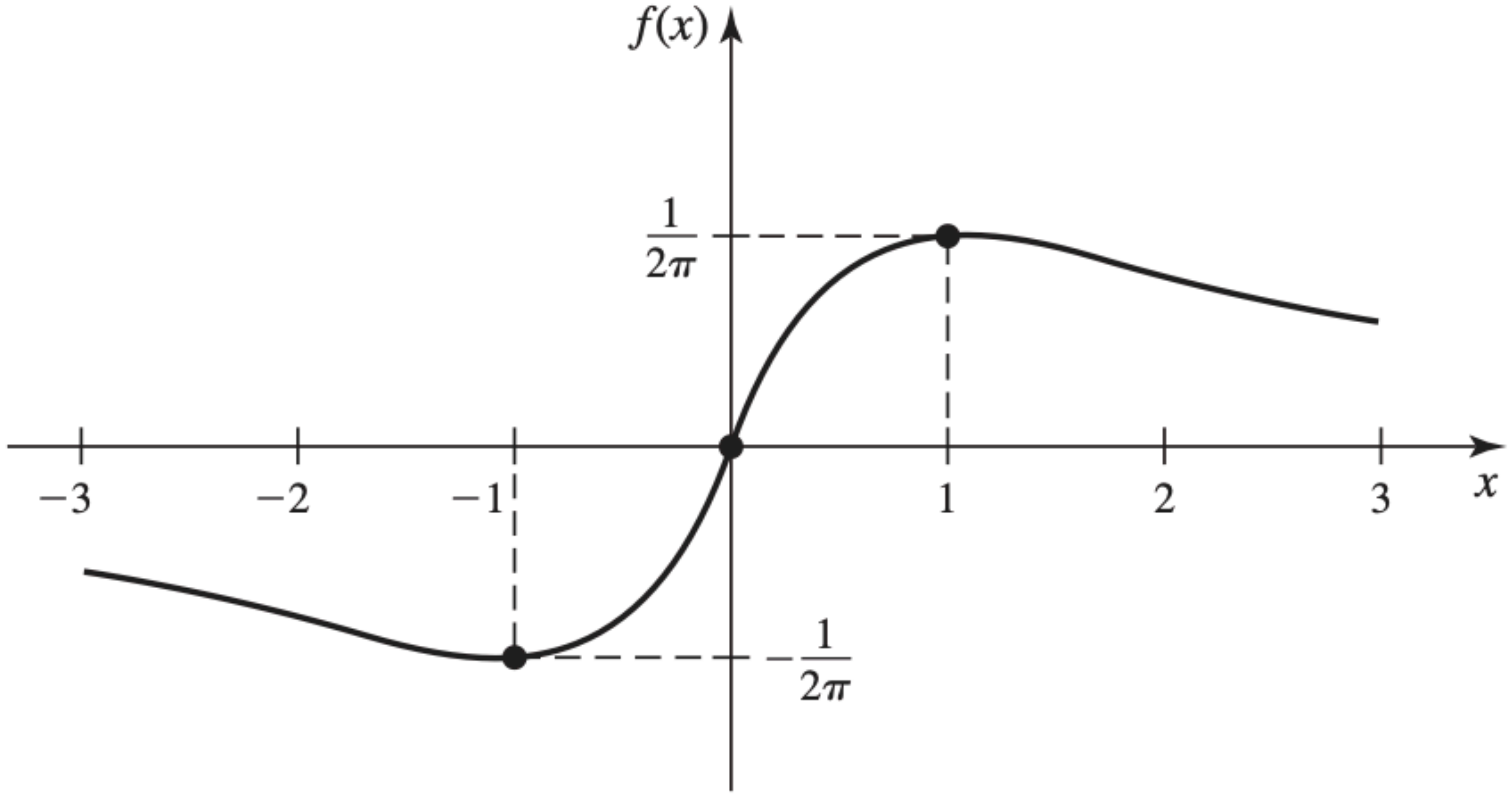


Figure 4.3 The p.d.f. of a Cauchy distribution.

Figure 4.4 The curve $y = xf(x)$ for the Cauchy distribution.



The Expectation of a Function

Example 4.1.10

Failure Rate and Time to Failure. Suppose that appliances manufactured by a particular company fail at a rate of X per year, where X is currently unknown and hence is a random variable. If we are interested in predicting how long such an appliance will last before failure, we might use the mean of $1/X$. How can we calculate the mean of $Y = 1/X$? ◄

Functions of a Single Random Variable If X is a random variable for which the p.d.f. is f, then the expectation of each real-valued function $r(X)$ can be found by applying the definition of expectation to the distribution of $r(X)$ as follows: Let $Y = r(X)$, determine the probability distribution of Y, and then determine $E(Y)$ by applying either Eq. (4.1.1) or Eq. (4.1.4). For example, suppose that Y has a continuous distribution with the p.d.f. g. Then

$$E[r(X)] = E(Y) = \int_{-\infty}^{\infty} yg(y)\, dy, \tag{4.1.8}$$

if the expectation exists.

Example 4.1.11 **Failure Rate and Time to Failure.** In Example 4.1.10, suppose that the p.d.f. of X is

$$f(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $r(x) = 1/x$. we can find the p.d.f. of $Y = r(X)$ as

$$g(y) = \begin{cases} 3y^{-4} & \text{if } y > 1, \\ 0 & \text{otherwise.} \end{cases}$$

The mean of Y is then

$$E(Y) = \int_0^\infty y3y^{-4}dy = \frac{3}{2}.$$ ◀

Although the method of Example 4.1.11 can be used to find the mean of a continuous random variable, it is not actually necessary to determine the p.d.f. of $r(X)$ in order to calculate the expectation $E[r(X)]$. In fact, it can be shown that the value of $E[r(X)]$ can always be calculated directly using the following result.

Theorem 4.1.1 **Law of the Unconscious Statistician.** Let X be a random variable, and let r be a real-valued function of a real variable. If X has a continuous distribution, then

$$E[r(X)] = \int_{-\infty}^{\infty} r(x) f(x)\, dx, \tag{4.1.9}$$

if the mean exists. If X has a discrete distribution, then

$$E[r(X)] = \sum_{\text{All } x} r(x) f(x), \tag{4.1.10}$$

if the mean exists.

Example 4.1.12 **Failure Rate and Time to Failure.** In Example 4.1.11, we can apply Theorem 4.1.1 to find

$$E(Y) = \int_0^1 \frac{1}{x} 3x^2 dx = \frac{3}{2},$$

the same result we got in Example 4.1.11. ◀

Example 4.1.13 **Determining the Expectation of $X^{1/2}$.** Suppose that the p.d.f. of X is as given in Example 4.1.6 and that $Y = X^{1/2}$. Then, by Eq. (4.1.9),

$$E(Y) = \int_0^1 x^{1/2}(2x)\,dx = 2\int_0^1 x^{3/2}\,dx = \frac{4}{5}.$$ ◀

Functions of Several Random Variables

Theorem 4.1.2 **Law of the Unconscious Statistician.** Suppose that $X_1, \ldots, X_n$ are random variables with the joint p.d.f. $f(x_1, \ldots, x_n)$. Let r be a real-valued function of n real variables, and suppose that $Y = r(X_1, \ldots, X_n)$. Then $E(Y)$ can be determined directly from the relation

$$E(Y) = \int \underset{R^n}{\cdots} \int r(x_1, \ldots, x_n) f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n,$$

if the mean exists. Similarly, if $X_1, \ldots, X_n$ have a discrete joint distribution with p.f. $f(x_1, \ldots, x_n)$, the mean of $Y = r(X_1, \ldots, X_n)$ is

$$E(Y) = \sum_{\text{All } x_1, \ldots, x_n} r(x_1, \ldots, x_n) f(x_1, \ldots, x_n),$$

if the mean exists. ■

Example 4.1.16

Determining the Expectation of a Function of Two Variables. Suppose that a point (X, Y) is chosen at random from the square S containing all points (x, y) such that $0 \le x \le 1$ and $0 \le y \le 1$. We shall determine the expected value of $X^2 + Y^2$.

Since X and Y have the uniform distribution over the square S, and since the area of S is 1, the joint p.d.f. of X and Y is

$$f(x, y) = \begin{cases} 1 & \text{for } (x, y) \in S, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} E(X^2 + Y^2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + y^2) f(x, y) \, dx \, dy \\ &= \int_0^1 \int_0^1 (x^2 + y^2) \, dx \, dy = \frac{2}{3}. \end{aligned}$$

◀

Summary

The expectation, expected value, or mean of a random variable is a summary of its distribution. If the probability distribution is thought of as a distribution of mass along the real line, then the mean is the center of mass. The mean of a function $\mathbf{r}$ of a random variable X can be calculated directly from the distribution of X without first finding the distribution of $\mathbf{r}(\mathbf{X})$. Similarly, the mean of a function of a random vector $\boldsymbol{X}$ can be calculated directly from the distribution of $\boldsymbol{X}$.

Properties of Expectations

Basic Theorems

Suppose that X is a random variable for which the expectation $E(X)$ exists. We shall present several results pertaining to the basic properties of expectations.

Theorem 4.2.1 Linear Function. If $Y = aX + b$, where a and b are finite constants, then

$$E(Y) = aE(X) + b.$$

Proof We first shall assume, for convenience, that X has a continuous distribution for which the p.d.f. is f. Then

$$\begin{aligned} E(Y) = E(aX + b) &= \int_{-\infty}^{\infty} (ax + b) f(x)\, dx \\ &= a \int_{-\infty}^{\infty} x f(x)\, dx + b \int_{-\infty}^{\infty} f(x)\, dx \\ &= aE(X) + b. \end{aligned}$$

A similar proof can be given for a discrete distribution. ■

Theorem 4.2.2 If there exists a constant such that $\Pr(X \geq a) = 1$, then $E(X) \geq a$. If there exists a constant b such that $\Pr(X \leq b) = 1$, then $E(X) \leq b$.

Proof We shall assume again, for convenience, that X has a continuous distribution for which the p.d.f. is f, and we shall suppose first that $\Pr(X \geq a) = 1$. Because X is bounded below, the second integral in (4.1.5) is finite. Then

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x)\,dx = \int_{a}^{\infty} x f(x)\,dx \\
&\geq \int_{a}^{\infty} a f(x)\,dx = a \Pr(X \geq a) = a.
\end{aligned}
$$

The proof of the other part of the theorem and the proof for a discrete distribution are similar. ■

Theorem 4.2.3 Suppose that $E(X) = a$ and that either $\Pr(X \geq a) = 1$ or $\Pr(X \leq a) = 1$. Then $\Pr(X = a) = 1$.

Proof We shall provide a proof for the case in which X has a discrete distribution and $\Pr(X \geq a) = 1$. The other cases are similar. Let $x_1, x_2, \ldots$ include every value $x > a$ such that $\Pr(X = x) > 0$, if any. Let $p_0 = \Pr(X = a)$. Then,

$$E(X) = p_0 a + \sum_{j=1}^{\infty} x_j \Pr(X = x_j). \tag{4.2.1}$$

Each x_j in the sum on the right side of Eq. (4.2.1) is greater than a. If we replace all of the x_j's by a, the sum can't get larger, and hence

$$E(X) \geq p_0 a + \sum_{j=1}^{\infty} a \Pr(X = x_j) = a. \tag{4.2.2}$$

Furthermore, the inequality in Eq. (4.2.2) will be strict if there is even one $x > a$ with $\Pr(X = x) > 0$. This contradicts $E(X) = a$. Hence, there can be no $x > a$ such that $\Pr(X = x) > 0$. ■

Theorem 4.2.4 If $X_1, \ldots, X_n$ are n random variables such that each expectation $E(X_i)$ is finite $(i = 1, \ldots, n)$, then

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n).$$

Proof We shall first assume that $n = 2$ and also, for convenience, that X_1 and X_2 have a continuous joint distribution for which the joint p.d.f. is f. Then

$$\begin{aligned}
E(X_1 + X_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1 + x_2) f(x_1, x_2)\, dx_1\, dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2)\, dx_1\, dx_2 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1, x_2)\, dx_1\, dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2)\, dx_2\, dx_1 + \int_{-\infty}^{\infty} x_2 f_2(x_2)\, dx_2 \\
&= \int_{-\infty}^{\infty} x_1 f_1(x_1)\, dx_1 + \int_{-\infty}^{\infty} x_2 f_2(x_2)\, dx_2 \\
&= E(X_1) + E(X_2),
\end{aligned}$$

where f_1 and f_2 are the marginal p.d.f.'s of X_1 and X_2. The proof for a discrete distribution is similar. Finally, the theorem can be established for each positive integer n by an induction argument. ■

Sampling without Replacement. Suppose that a box contains red balls and blue balls and that the proportion of red balls in the box is p $(0 \le p \le 1)$. Suppose that n balls are selected from the box at random *without replacement*, and let X denote the number of red balls that are selected. We shall determine the value of $E(X)$.

Sampling with Replacement. Suppose again that in a box containing red balls and blue balls, the proportion of red balls is p $(0 \le p \le 1)$. Suppose now, however, that a random sample of n balls is selected from the box *with replacement*. If X denotes the number of red balls in the sample, then X has the binomial distribution with parameters n and p, as described in Sec. 3.1. We shall now determine the value of $E(X)$.

Expected Number of Matches. Suppose that a person types n letters, types the addresses on n envelopes, and then places each letter in an envelope in a random manner. Let X be the number of letters that are placed in the correct envelopes. We shall find the mean of X.

For $i = 1, \ldots, n$, let $X_i = 1$ if the ith letter is placed in the correct envelope, and let $X_i = 0$ otherwise. Then, for $i = 1, \ldots, n$,

$$\Pr(X_i = 1) = \frac{1}{n} \quad \text{and} \quad \Pr(X_i = 0) = 1 - \frac{1}{n}.$$

Therefore,

$$E(X_i) = \frac{1}{n} \quad \text{for } i = 1, \ldots, n.$$

Since $X = X_1 + \cdots + X_n$, it follows that

$$\begin{aligned} E(X) &= E(X_1) + \cdots + E(X_n) \\ &= \frac{1}{n} + \cdots + \frac{1}{n} = 1. \end{aligned}$$

Thus, the expected value of the number of correct matches of letters and envelopes is 1, regardless of the value of n. ◀

Expectation of a Product of Independent Random Variables

Theorem 4.2.6 If $X_1, \ldots, X_n$ are n independent random variables such that each expectation $E(X_i)$ is finite ($i = 1, \ldots, n$), then

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i).$$

Proof We shall again assume, for convenience, that $X_1, \ldots, X_n$ have a continuous joint distribution for which the joint p.d.f. is f. Also, we shall let f_i denote the marginal p.d.f. of X_i ($i = 1, \ldots, n$). Then, since the variables $X_1, \ldots, X_n$ are independent, it follows that at every point $(x_1, \ldots, x_n) \in R^n$,

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i).$$

Therefore,

$$\begin{aligned}
E\left(\prod_{i=1}^{n} X_i\right) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} x_i\right) f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} x_i f_i(x_i)\right] dx_1 \ldots dx_n \\
&= \prod_{i=1}^{n} \int_{-\infty}^{\infty} x_i f_i(x_i)\, dx_i = \prod_{i=1}^{n} E(X_i).
\end{aligned}$$

The proof for a discrete distribution is similar. ■

Example 4.2.7 **Calculating the Expectation of a Combination of Random Variables.** Suppose that X_1, X_2, and X_3 are independent random variables such that $E(X_i) = 0$ and $E(X_i^2) = 1$ for $i = 1, 2, 3$. We shall determine the value of $E[X_1^2(X_2 - 4X_3)^2]$.

Since X_1, X_2, and X_3 are independent, it follows that the two random variables X_1^2 and $(X_2 - 4X_3)^2$ are also independent. Therefore,

$$\begin{aligned} E[X_1^2(X_2 - 4X_3)^2] &= E(X_1^2)E[(X_2 - 4X_3)^2] \\ &= E(X_2^2 - 8X_2X_3 + 16X_3^2) \\ &= E(X_2^2) - 8E(X_2X_3) + 16E(X_3^2) \\ &= 1 - 8E(X_2)E(X_3) + 16 \\ &= 17. \end{aligned}$$

◀

Formulas of mathematical expectation and dispersion of a continuous random variable X have the following form:

$$\boxed{a = M(X) = \int_{-\infty}^{+\infty} x\varphi(x)dx}$$ (if the integral converges absolutely)

$$\boxed{D(X) = \int_{-\infty}^{+\infty} (x-a)^2 \varphi(x)dx}$$ (if the integral converges).

Using $D(X) = M(X^2) - [M(X)]^2$, we have $D(X) = M(X^2) - a^2$ or

$$\boxed{D(X) = \int_{-\infty}^{+\infty} x^2\varphi(x)dx - a^2}$$