# 3

# Expectation

## 3.1 INTRODUCTION

We begin here the study of the long-run convergence properties of situations involving a very large number of independent repetitions of a random experiment. As an introductory example, suppose that we observe the length of a telephone call made from a specific phone booth at a given time of the day, say, the first call after 12 o'clock noon. Suppose that we repeat the experiment independently $n$ times, where $n$ is very large, and record the cost of each call (which is determined by its length). If we take the arithmetic average of the costs, that is, add the total cost of all $n$ calls and then divide by $n$, we expect physically that the arithmetic average will converge in some sense to a number that we should interpret as the long-run average cost of a call. We shall try first to pin down the notion of average more precisely.

Assume that the cost $R_2$ of a call in terms of its length $R_1$ is as follows.

| | |
|---|---|
| If $0 \leq R_1 \leq 3$ (minutes) | $R_2 = 10$ (cents) |
| If $3 < R_1 \leq 6$ | $R_2 = 20$ |
| If $6 < R_1 \leq 9$ | $R_2 = 30$ |

(Assume for simplicity that the telephone is automatically disconnected after 9 minutes.)

Thus $R_2$ takes on three possible values, 10, 20, and 30; say $P\{R_2 = 10\} = .6$, $P\{R_2 = 20\} = .25$, $P\{R_2 = 30\} = .15$. If we observe $N$ calls, where $N$ is very large, then, roughly, $\{R_2 = 10\}$ will occur $.6N$ times; the total cost of calls of this type is $10(.6N) = 6N$. $\{R_2 = 20\}$ will occur approximately $.25N$ times, giving rise to a total cost of $20(.25N) = 5N$. $\{R_2 = 30\}$ will occur approximately $.15N$ times, producing a total cost of $30\ (.15N) = 4.5N$. The total cost of all calls is $6N + 5N + 4.5N = 15.5N$, or 15.5 cents per call on the average.

Observe how we have computed the average.

$$\frac{10(.6N) + 20(.25N) + 30(.15N)}{N} = 10(.6) + 20(.25) + 30(.15)$$
$$= \sum_y yP\{R_2 = y\}$$

Thus we are taking a *weighted average* of the possible values of $R_2$, where the weights are the probabilities of $R_2$ assuming those values. This suggests the following definition.

Let $R$ be a *simple* random variable, that is, a discrete random variable taking on only *finitely* many possible values. Define the *expectation* [also called the *expected value*, *average value*, *mean value*, or *mean*] of $R$ as

$$E(R) = \sum_x xP\{R = x\} \tag{3.1.1}$$

Since $R$ is simple, this is a finite sum and there are no convergence problems. In particular, if $R$ is identically constant, say $R = c$, then $E(R) = cP\{R = c\} = c$. For short,

$$E(c) = c \tag{3.1.2}$$

Note that if $R$ takes the values $x_1, \ldots, x_n$, each with probability $1/n$, then $E(R) = (x_1 + \cdots + x_n)/n$, as we would expect intuitively. In this case each $x_i$ is given the same weight, namely, $1/n$.

We now have the problem of extending the definition to more general random variables. If $R$ is an arbitrary discrete random variable, the natural choice for $E(R)$ is again $\sum_x xP\{R = x\}$, provided that the sum makes sense. (Theorem 1 will make this precise.)

Similarly, let $R_1$ be discrete and $R_2 = g(R_1)$. Since $R_2$ is also discrete, we have $E(R_2) = \sum_y yP\{R_2 = y\}$. However, if $x_1, x_2, \ldots$ are the values of $R_1$, then with probability $p_{R_1}(x_i)$ we have $R_1 = x_i$, hence $R_2 = g(x_i)$. Thus if our definition of expectation is sound, we should have the following alternate expression for $E(R_2)$:

$$E(R_2) = E[g(R_1)] = \sum_i g(x_i)p_{R_1}(x_i) \tag{3.1.3}$$

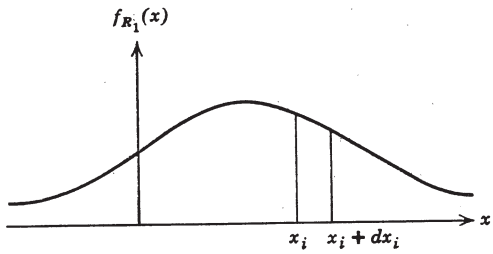


FIGURE 3.1.1

again a weighted average of possible values of $R_2$, but expressed in terms of the probability function of $R_1$.

If $R_1$ is absolutely continuous, this approach breaks down completely, since $P\{R_1 = x\} = 0$ for all $x$. However, we may get some idea as to how to compute $E(R_2) = E[g(R_1)]$ when $R_1$ is absolutely continuous, by making a discrete approximation. If we split the real line into intervals $(x_i, x_i + dx_i]$, then, roughly, the probability that $x_i < R_1 \leq x_i + dx_i$ is $f_{R_1}(x_i)\, dx_i$ (see Figure 3.1.1). If $R_1$ falls into this interval, $g(R_1)$ is approximately $g(x_i)$, at least if $g$ is continuous. Thus an approximation to $E(R_2)$ should be

$$\sum_i g(x_i) f_R(x_i)\, dx_i$$

which suggests that if a general definition of expectation is formulated properly, and $R_1$ is absolutely continuous,

$$E[g(R_1)] = \int_{-\infty}^{\infty} g(x) f_{R_1}(x)\, dx \tag{3.1.4}$$

In the telephone call example above, if $R_1$ has density $f_1$, we obtain

$$E(R_2) = \int_{-\infty}^{\infty} g(x) f_1(x)\, dx$$

where

$$\begin{aligned} R_2 = g(R_1) &= 10 \qquad \text{if } 0 \leq R_1 \leq 3 \\ &= 20 \qquad \text{if } 3 < R_1 \leq 6 \\ &= 30 \qquad \text{if } 6 < R_1 \leq 9 \end{aligned}$$

Thus

$$\begin{aligned} E(R_2) &= 10 \int_0^3 f_1(x)\, dx + 20 \int_3^6 f_1(x)\, dx + 30 \int_6^9 f_1(x)\, dx \\ &= 10(.6) + 20(.25) + 30(.15) \qquad \text{as before} \end{aligned}$$

If we have an $n$-dimensional situation, for instance $R_0 = g(R_1, \ldots, R_n)$,

the preceding formulas generalize in a natural way. If $R_1, \ldots, R_n$ are discrete,

$$E[g(R_1, \ldots, R_n)] = \sum_{x_1, \ldots, x_n} g(x_1, \ldots, x_n) P\{R_1 = x_1, \ldots, R_n = x_n\} \quad (3.1.5)$$

If $(R_1, \ldots, R_n)$ is absolutely continuous with density $f_{12\ldots n}$,

$$E[g(R_1, \ldots, R_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{12\ldots n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \quad (3.1.6)$$

We shall outline very briefly a general definition of expectation that includes all the previous special cases.

If $R$ is a simple random variable on $(\Omega, \mathcal{F}, P)$, we define

$$E(R) = \sum_x x P\{R = x\}$$

just as above. Now let $R$ be a nonnegative random variable. We approximate $R$ by simple random variables as follows.

Define

$$R_n(\omega) = \frac{k-1}{2^n} \quad \text{if } \frac{k-1}{2^n} \leq R(\omega) < \frac{k}{2^n}, \qquad k = 1, 2, \ldots, n2^n$$

and let

$$R_n(\omega) = n \quad \text{if } R(\omega) \geq n$$

(see Figure 3.1.2 for an illustration with $n = 2$).

For any fixed $\omega$, eventually $R(\omega) < n$, so that $0 \leq R(\omega) - R_n(\omega) < 2^{-n}$. Thus $R_n(\omega) \to R(\omega)$. In fact $R_n(\omega) \leq R_{n+1}(\omega)$ for all $n$, $\omega$. For example, if $3/4 \leq R(\omega) < 7/8$, then $R_2(\omega) = R_3(\omega) = 3/4$; if $7/8 \leq R(\omega) < 1$, then $R_2(\omega) = 3/4$, $R_3(\omega) = 7/8$. In general, if $R(\omega)$ lies in the lower half of the interval $[(k-1)/2^n, k/2^n)$, then $R_n(\omega) = R_{n+1}(\omega)$; if $R(\omega)$ lies in the upper half, $R_n(\omega) < R_{n+1}(\omega)$.

Thus we have constructed a sequence of nonnegative simple functions $R_n$ converging monotonically up to $R$. We have already defined $E(R_n)$, and since $R_n \leq R_{n+1}$ we have $E(R_n) \leq E(R_{n+1})$. We define

$$E(R) = \lim_{n \to \infty} E(R_n) \qquad \text{(this may be } +\infty)$$

It is possible to show that if $\{R_n'\}$ is any other sequence of nonnegative simple functions converging monotonically up to $R$,

$$\lim_{n \to \infty} E(R_n') = \lim_{n \to \infty} E(R_n)$$

and thus $E(R)$ is well defined.

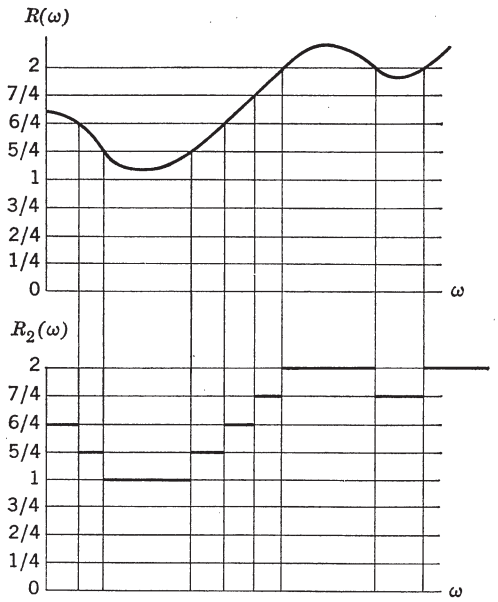


FIGURE 3.1.2 Approximation of a Nonnegative Random Variable by Simple Random Variables.

Finally, if $R$ is an arbitrary random variable, let $R^+ = \max(R, 0)$, $R^- = \max(-R, 0)$; that is,

$$R^+(\omega) = \begin{cases} R(\omega) & \text{if } R(\omega) \geq 0; \\ 0 & \text{if } R(\omega) < 0 \end{cases}$$

$$R^-(\omega) = \begin{cases} -R(\omega) & \text{if } R(\omega) \leq 0; \\ 0 & \text{if } R(\omega) > 0 \end{cases}$$

$R^+$ and $R^-$ are called the *positive* and *negative* parts of $R$ (see Figure 3.1.3). It follows that $R = R^+ - R^-$ (and $|R| = R^+ + R^-$), and we define $E(R) = E(R^+) - E(R^-)$ if this is not of the form $+\infty - \infty$; if it is, we say that the expectation does not exist. Note that $E(R)$ is finite if and only if $E(R^+)$ and $E(R^-)$ are both finite. Since it can be shown that $E(|R|) = E(R^+) + E(R^-)$, it follows that

$$E(R) \text{ is finite if and only if } E(|R|) \text{ is finite} \tag{3.1.7}$$

The expectation of a nonnegative random variable always exists; it may be $+\infty$.

The following results may be proved.

Let $R_1, R_2, \ldots, R_n$ be random variables on $(\Omega, \mathcal{F}, P)$, and let $R_0 = g(R_1, \ldots, R_n)$, where $g$ is a function from $E^n$ to $E^1$. Assume that $g$ has the property that $g^{-1}(B)$ is a Borel subset of $E^n$ whenever $B$ is a Borel subset of $E^1$. Then, as we indicated in Section 2.7, $R_0$ is a random variable.

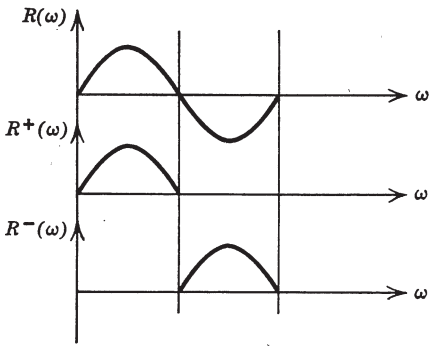


FIGURE 3.1.3 Positive and Negative Parts of a Random Variable.

***Theorem 1.*** *If $R_1, \ldots, R_n$ are discrete, then*

$$E[g(R_1, \ldots, R_n)] = \sum_{x_1, \ldots, x_n} g(x_1, \ldots, x_n) P\{R_1 = x_1, \ldots, R_n = x_n\}$$

*if $g(x_1, \ldots, x_n) \geq 0$ for all $x_1, \ldots, x_n$, or if the series on the right is absolutely convergent.*

***Theorem 2.*** *If $(R_1, \ldots, R_n)$ is absolutely continuous with density $f_{12\ldots n}$, then*

$$E[g(R_1, \ldots, R_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{12\ldots n}(x_1, \ldots, x_n)\, dx_1, \ldots, dx_n$$

*if $g(x_1, \ldots, x_n) \geq 0$ for all $x_1, \ldots, x_n$, or if the integral on the right is absolutely convergent.*

We shall look at examples that are neither discrete nor absolutely continuous in Chapter 4.

Notice that it is quite possible for the expectation to exist and be infinite, or not to exist at all. For example, let

$$f_R(x) = \frac{1}{x^2}, \qquad x \geq 1; \qquad f_R(x) = 0, \qquad x < 1$$

(see Figure 3.1.4a). Then

$$E(R) = \int_{-\infty}^{\infty} x f_R(x)\, dx = \int_1^{\infty} x \frac{1}{x^2}\, dx = \infty$$

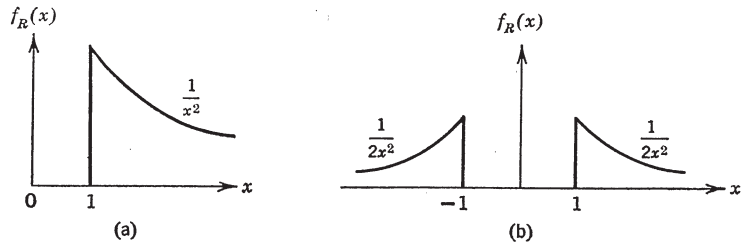


FIGURE 3.1.4 (a) $E(R) = \infty$. (b) $E(R)$ Does Not Exist.

As another example, let $f_R(x) = 1/2x^2$, $|x| \geq 1$; $f_R(x) = 0$, $|x| < 1$ (Figure 3.1.4b). Then (see Figure 3.1.5)

$$E(R^+) = \int_{-\infty}^{\infty} x^+ f_R(x)\,dx = \int_0^{\infty} x f_R(x)\,dx = \tfrac{1}{2}\int_0^{\infty} x\,\frac{1}{x^2}\,dx = \infty$$

$$E(R^-) = \int_{-\infty}^{\infty} x^- f_R(x)\,dx = \int_{-\infty}^{0} -x f_R(x)\,dx = \tfrac{1}{2}\int_{-\infty}^{-1} -x\,\frac{1}{x^2}\,dx$$

$$= \tfrac{1}{2}\int_1^{\infty} \frac{1}{x}\,dx = \infty$$

Thus $E(R)$ does not exist.

Finally it can be shown that if two random variables $R_1$ and $R_2$ are "essentially" equal, that is, if $P\{\omega\colon R_1(\omega) \neq R_2(\omega)\} = 0$, then $E(R_1) = E(R_2)$ if the expectations exist.

REMARK. Theorem 1 fails if the series on the right is conditionally but not absolutely convergent. For example, let $P\{R_1 = n\} = (1/2)^n$, $n = 1, 2, \ldots,$ and $R_2 = g(R_1)$, where $g(n) = (-1)^{n+1}2^n/n$. If $R_1(\omega) = n$, $n$ odd, then $R_2(\omega) = 2^n/n$; hence $R_2^+(\omega) = g(n) = 2^n/n$, $R_2^-(\omega) = 0$. If $R_1(\omega) = n$, $n$ even, then $R_2(\omega) = -2^n/n$; hence $R_2^+(\omega) = 0$, $R_2^-(\omega) = -g(n) = 2^n/n$. Therefore, by the nonnegative case of Theorem 1,

$$E(R_2^+) = \sum_{n \text{ odd}} g(n)P\{R_1 = n\} = 1 + \tfrac{1}{3} + \tfrac{1}{5} + \cdots = \infty$$

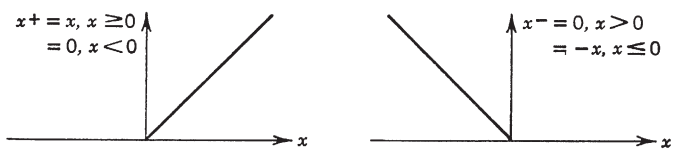


FIGURE 3.1.5

and

$$E(R_2^-) = \sum_{n \text{ even}} -g(n)P\{R_1 = n\} = \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{6} + \cdots = \infty$$

Hence $E(R_2)$ does not exist, although

$$\sum_{n=1}^{\infty} g(n)P\{R_1 = n\} = 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \cdots$$

is conditionally convergent. From an intuitive standpoint, the expectation should not change if the series is rearranged; a conditionally but not absolutely convergent series will not have this property.

## 3.2 TERMINOLOGY AND EXAMPLES

If $R$ is a random variable on a given probability space, the $k$th *moment* of $R$ ($k > 0$, not necessarily an integer) is defined by

$$\alpha_k = E(R^k) \qquad \text{if the expectation exists}$$

Thus

$$\alpha_k = \sum_x x^k p_R(x) \qquad \text{if } R \text{ is discrete}$$

$$= \int_{-\infty}^{\infty} x^k f_R(x)\, dx \qquad \text{if } R \text{ is absolutely continuous}$$

$\alpha_1$ is simply $E(R)$, the expectation of $R$, often written as $m$ and called the *mean* of $R$. If $R$ has density $f_R$, $m$ may be regarded as the abscissa of the centroid of the region in the plane between the $x$-axis and the graph of $f_R$ (see Figure 3.2.1). To see this, notice that the total moment of the region

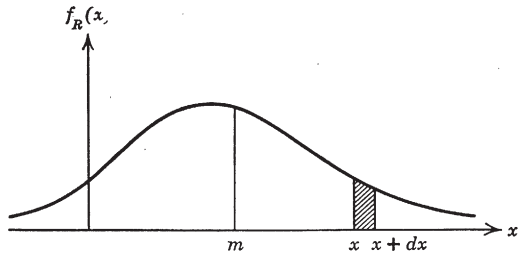


FIGURE 3.2.1 Geometric Interpretation of $E(R)$. The "Strip" Between $x$ and $x + dx$ Contributes $(x - m)f_R(x)\, dx$ to the Moment about $x = m$.

about the line $x = m$ is

$$\int_{-\infty}^{\infty} (x - m) f_R(x)\, dx = m - m = 0$$

The expectation of $R$ is a *measure of central tendency* in the sense that the arithmetic average of $n$ independent observations of $R$ converges (in a sense yet to be made precise) to $E(R)$.

The $k$th *central moment* of $R$ ($k > 0$) is defined by

$$\begin{aligned}
\beta_k &= E[(R - m)^k] \qquad \text{if } m \text{ is finite and the expectation exists} \\
&= \sum_x (x - m)^k p_R(x) \qquad \text{if } R \text{ is discrete} \\
&= \int_{-\infty}^{\infty} (x - m)^k f_R(x)\, dx \qquad \text{if } R \text{ is absolutely continuous}
\end{aligned}$$

Notice that $\beta_1 = E(R - m) = m - m = 0$.

$\beta_2 = E[(R - m)^2]$ is called the *variance* of $R$, written $\sigma^2$, $\sigma^2(R)$, or Var $R$. $\sigma$ (the positive square root of $\beta_2$) is called the *standard deviation* of $R$. Note that if $R$ has finite mean, then, since $(R - m)^2 \geq 0$, Var $R$ always exists; it may be infinite.

If $R$ has density $f_R$, the variance of $R$ may be regarded as the moment of inertia of the region in the plane between the $x$-axis and the graph of $f_R$, about the axis $x = m$.

The variance may be interpreted as a *measure of dispersion*. A large variance corresponds to a high probability that $R$ will fall far from its mean, while a small variance indicates that $R$ is likely to be close to its mean (see Figure 3.2.2). We shall make a quantititative statement to this effect (Chebyshev's inequality) in Section 3.7.

▶ **Example 1.** Consider the *normal density function*

$$f_R(x) = \frac{1}{\sqrt{2\pi}\, b} e^{-(x-a)^2/2b^2}, \qquad b > 0, \; a \text{ real}$$

Since $f_R$ is symmetrical about $x = a$, the centroid of the area under $f_R$ has abscissa $a$, so that $E(R) = a$. We compute the variance of $R$.

$$\sigma^2 = \int_{-\infty}^{\infty} \frac{(x - a)^2}{\sqrt{2\pi}\, b} e^{-(x-a)^2/2b^2} dx$$

Let $y = (x - a)/\sqrt{2}\, b$. We obtain

$$\sigma^2 = \int_{-\infty}^{\infty} \frac{2b^2}{\sqrt{2\pi}\, b} y^2 e^{-y^2} \sqrt{2}\, b\, dy = \frac{2b^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2} dy$$

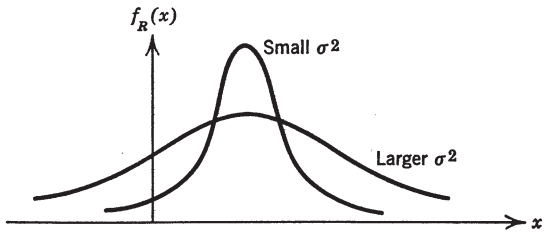


FIGURE 3.2.2

Now, by (2.8.2),

$$\sqrt{\pi} = \int_{-\infty}^{\infty} e^{-y^2}\, d(y)$$

Integrate by parts to obtain

$$\sqrt{\pi} = ye^{-y^2}]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} - 2y^2\, e^{-y^2}\, dy$$

It follows that

$$\int_{-\infty}^{\infty} y^2 e^{-y^2}\, dy = \tfrac{1}{2}\sqrt{\pi}$$

Hence $\sigma^2 = b^2$. Thus we may write

$$f_R(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-m)^2/2\sigma^2} \tag{3.2.1}$$

In this case the mean and variance determine the density completely.

If $R$ has the normal density with mean $m$ and variance $\sigma^2$, we sometimes write "$R$ is normal $(m, \sigma^2)$" for short. ◀

Before looking at the next example, it will be convenient to introduce the *gamma function*, defined by

$$\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x}\, dx, \qquad r > 0 \tag{3.2.2}$$

Integrating by parts, we have

$$\Gamma(r) = \int_0^{\infty} e^{-x}\, d\left(\frac{x^r}{r}\right) = \frac{x^r e^{-x}}{r}\Big]_0^{\infty} + \int_0^{\infty} \frac{x^r}{r}\, e^{-x}\, dx$$

$$= \int_0^{\infty} \frac{x^r}{r}\, e^{-x}\, dx = \frac{\Gamma(r+1)}{r}$$

Thus

$$\Gamma(r+1) = r\Gamma(r) \tag{3.2.3}$$

Since

$$\Gamma(1) = \int_0^\infty e^{-x}\,dx = 1$$

we have

$$\Gamma(2) = 1\Gamma(1) = 1, \quad \Gamma(3) = 2\Gamma(2) = 2 \cdot 1, \quad \Gamma(4) = 3\Gamma(3) = 3 \cdot 2 \cdot 1 = 3!$$

and

$$\Gamma(n+1) = n!, \qquad n = 0, 1, \ldots \tag{3.2.4}$$

We also need $\Gamma(1/2)$.

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty x^{-1/2}e^{-x}\,dx$$

Let $x = y^2$ to obtain

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty \frac{1}{y}\,e^{-y^2}2y\,dy = 2\int_0^\infty e^{-y^2}\,dy = \int_{-\infty}^\infty e^{-y^2}\,dy$$

By (2.8.2), we have

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi} \tag{3.2.5}$$

▶ **Example 2.** Let $R_1$ be absolutely continuous with density $f_1(x) = e^{-x}$, $x \geq 0$; $f_1(x) = 0$, $x < 0$. Let $R_2 = R_1^2$. We may compute $E(R_2)$ in two ways.

1. $E(R_2) = E(R_1^2) = \int_{-\infty}^\infty x^2 f_1(x)\,dx = \int_0^\infty x^2 e^{-x}\,dx = \Gamma(3) = 2$ by (3.2.4)

2. We may find the density of $R_2$ by the technique of Section 2.4 (see Figure 3.2.3). We have

$$f_2(y) = f_1(\sqrt{y})\frac{d}{dy}\sqrt{y} = \frac{e^{-\sqrt{y}}}{2\sqrt{y}}, \qquad y > 0$$
$$= 0, \qquad y < 0$$

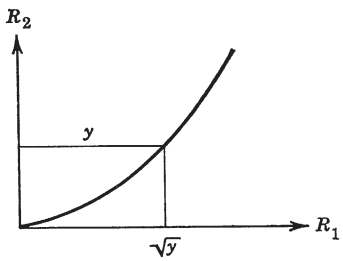


FIGURE 3.2.3 Computation of Density of $R_2$.

Then

$$E(R_2) = \int_{-\infty}^{\infty} y f_2(y)\, dy$$
$$= \int_0^{\infty} y \frac{e^{-\sqrt{y}}}{2\sqrt{y}}\, dy$$
$$= (\text{with } y = x^2) \int_0^{\infty} x^2 e^{-x}\, dx = 2$$

as before.

Notice that both methods must give the same answer by Theorem 2 of Section 3.1. For $R_2(\omega) = (R_1(\omega))^2$; applying the theorem with $g(R_1) = R_1^2$, we obtain

$$E(R_1^2) = \int_{-\infty}^{\infty} x^2 f_1(x)\, dx$$

Applying the theorem with $g(R_2) = R_2$, we have

$$E(R_2) = \int_{-\infty}^{\infty} y f_2(y)\, dy$$

Generally the first method is easier, since the computation of the density of $R_2$ is avoided. ◀

▶ **Example 3.** Let $R_1$ and $R_2$ be independent, each with density $f(x) = e^{-x}$, $x \geq 0$; $f(x) = 0$, $x < 0$. Let $R_3 = \max(R_1, R_2)$. We compute $E(R_3)$.

$$E(R_3) = E[g(R_1, R_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{12}(x, y)\, dx\, dy$$
$$= \int_0^{\infty}\int_0^{\infty} \max(x, y) e^{-x} e^{-y}\, dx\, dy$$

Now $\max(x, y) = x$ if $x \geq y$; $\max(x, y) = y$ if $x \leq y$ (see Figure 3.2.4). Thus

$$E(R_3) = \iint_A x e^{-x} e^{-y}\, dx\, dy + \iint_B y e^{-x} e^{-y}\, dx\, dy$$
$$= \int_{x=0}^{\infty} x e^{-x} \int_{y=0}^{x} e^{-y}\, dy\, dx + \int_{y=0}^{\infty} y e^{-y} \int_{x=0}^{y} e^{-x}\, dx\, dy$$

on $A$, $\max(x, y) = x$
on $B$, $\max(x, y) = y$

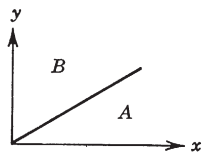


FIGURE 3.2.4

The two integrals are equal, since one may be obtained from the other by interchanging $x$ and $y$. Thus

$$E(R_3) = 2\int_0^\infty xe^{-x}\int_0^x e^{-y}\,dy\,dx = 2\int_0^\infty xe^{-x}(1 - e^{-x})\,dx$$
$$= 2\int_0^\infty xe^{-x}\,dx - 2\int_0^\infty \frac{z}{2}e^{-z}\frac{dz}{2} = \frac{3}{2}\Gamma(2) = \frac{3}{2} \blacktriangleleft$$

The moments and central moments of a random variable $R$, especially the mean and variance, give some information about the behavior of $R$. In many situations it may be difficult to compute the distribution function of $R$ explicitly, but the calculation of some of the moments may be easier. We shall examine some problems of this type in Section 3.5.

Another parameter that gives some information about a random variable $R$ is the *median* of $R$, defined when $F_R$ is continuous as a number $\mu$ (not necessarily unique) such that $F_R(\mu) = 1/2$ (see Figure 3.2.5a and b).

In general the median of a random variable $R$ is a number $\mu$ such that

$$F_R(\mu) = P\{R \leq \mu\} \geq \tfrac{1}{2}$$
$$F_R(\mu^-) = P\{R < \mu\} \leq \tfrac{1}{2}$$

(see Figure 3.2.5c).

Loosely speaking, $\mu$ is the halfway point of the distribution function of $R$.

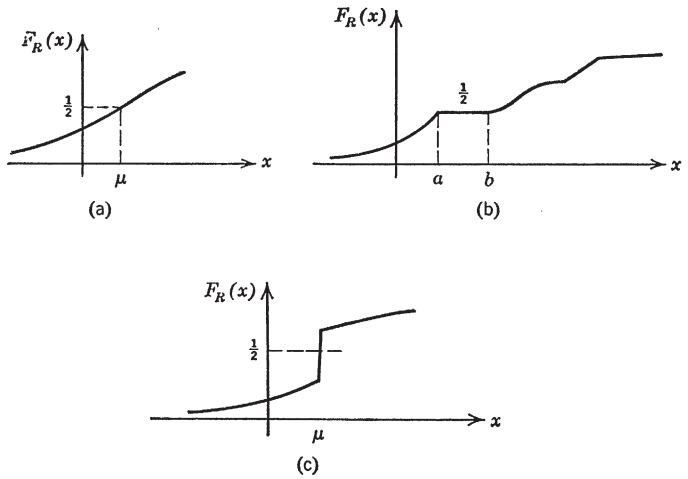


FIGURE 3.2.5 (a) $\mu$ is the Unique Median. (b) Any Number Between $a$ and $b$ is a Median. (c) $\mu$ is the Unique Median.

## PROBLEMS

**1.** Let $R$ be normally distributed with mean 0 and variance 1. Show that

$$\begin{aligned} E(R^n) &= 0, \qquad n \text{ odd} \\ &= (n-1)(n-3)\cdots(5)(3)(1), \qquad n \text{ even} \end{aligned}$$

**2.** Let $R_1$ have the exponential density $f_1(x) = e^{-x}$, $x \geq 0$; $f_1(x) = 0$, $x < 0$. Let $R_2 = g(R_1)$ be the largest integer $\leq R_1$ (if $0 \leq R_1 < 1$, $R_2 = 0$; if $1 \leq R_1 < 2$, $R_2 = 1$, and so on).
(a) Find $E(R_2)$ by computing $\int_{-\infty}^{\infty} g(x) f_1(x)\, dx$.
(b) Find $E(R_2)$ by evaluating the probability function of $R_2$ and then computing $\sum_y y p_{R_2}(y)$.

**3.** Let $R_1$ and $R_2$ be independent random variables, each with the exponential density $f(x) = e^{-x}$, $x \geq 0$; $f(x) = 0$, $x < 0$. Find the expectation of
(a) $R_1 R_2$
(b) $R_1 - R_2$
(c) $|R_1 - R_2|$

**4.** Let $R_1$ and $R_2$ be independent, each uniformly distributed between $-1$ and $+1$. Find $E[\max(R_1, R_2)]$.

**5.** Suppose that the density function for the length $R$ of a telephone call is

$$\begin{aligned} f(x) &= xe^{-x}, \qquad x \geq 0 \\ &= 0, \qquad x < 0 \end{aligned}$$

The cost of a call is

$$\begin{aligned} C(R) &= 2, \qquad 0 < R \leq 3 \\ &= 2 + 6(R-3), \qquad R > 3 \end{aligned}$$

Find the average cost of a call.

**6.** Two machines are put into service at $t = 0$, processing the same data. Let $R_i$ ($i = 1, 2$) be the time (in hours) at which machine $i$ breaks down. Assume that $R_1$ and $R_2$ are independent random variables, each having the exponential density function $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$; $f(x) = 0$, $x < 0$. Suppose that we start counting down time if and only if *both* machines are out of service. No repairs are allowed during the working day (which is $T$ hours long), but any machine that has failed during the day is assumed to be completely repaired by the time the next day begins. For example, if $T = 8$ and the machines fail at $t = 2$ and $t = 6$, the down time is 2 hours.
(a) Find the probability that at least one machine will fail during a working day.
(b) Find the average down time per day. (Leave the answer in the form of an integral.)

**7.** Show that if $R$ has the binomial distribution with parameters $n$ and $p$, that is, $R$ is the number of successes in $n$ Bernoulli trials with probability of success $p$ on

a given trial, then $E(R) = np$, as one should expect intuitively. HINT: in $E(R) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k}$, factor out $np$ and use the binomial theorem.

REMARK. In Section 3.5 we shall calculate the mean and variance of $R$ in an indirect but much more efficient way.

**8.** If $R$ has the Poisson distribution with parameter $\lambda$, show that

$$E[R(R-1)(R-2)\cdots(R-r+1)] = \lambda^r$$

Conclude that $E(R) = \text{Var } R = \lambda$.

## 3.3 PROPERTIES OF EXPECTATION

In this section we list several basic properties of the expectation of a random variable. A precise justification of these properties would require a detailed analysis of the general definition of $E(R)$ that we gave in Section 3.1; what we actually did there was to outline the construction of the abstract Lebesgue integral. Instead we shall give plausibility arguments or proofs in special cases.

1. Let $R_1, \ldots, R_n$ be random variables on a given probability space. Then

$$E(R_1 + \cdots + R_n) = E(R_1) + \cdots + E(R_n)$$

CAUTION. Recall that $E(R)$ can be $\pm\infty$, or not exist at all. The complete statement of property 1 is: If $E(R_i)$ exists for all $i = 1, 2, \ldots, n$, and $+\infty$ and $-\infty$ do not *both* appear in the sum $E(R_1) + \cdots + E(R_n)$ ($+\infty$ alone or $-\infty$ alone is allowed), then $E(R_1 + \cdots + R_n)$ exists and equals $E(R_1) + \cdots + E(R_n)$.

For example, suppose that $(R_1, R_2)$ has density $f_{12}$, and $R' = g(R_1, R_2)$, $R'' = h(R_1, R_2)$. Then

$$\begin{aligned} E(R' + R'') &= E[g(R_1, R_2) + h(R_1, R_2)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [g(x, y) + h(x, y)]f_{12}(x, y)\,dx\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f_{12}(x, y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y)f_{12}(x, y)\,dx\,dy \\ &= E(R') + E(R'') \end{aligned}$$

2. If $R$ is a random variable whose expectation exists, and $a$ is any real number, then $E(aR)$ exists and

$$E(aR) = aE(R)\dagger$$

For example, if $R_1$ has density $f_1$ and $R_2 = aR_1$, then

$$E(R_2) = \int_{-\infty}^{\infty} axf_1(x)\,dx = aE(R_1)$$

Basically, properties 1 and 2 say that the expectation is linear.

3. If $R_1 \leq R_2$, then $E(R_1) \leq E(R_2)$, assuming that both expectations exist. For example, if $R$ has density $f$, and $R_1 = g(R)$, $R_2 = h(R)$, and $g \leq h$, we have

$$E(R_1) = \int_{-\infty}^{\infty} g(x)f(x)\,dx \leq \int_{-\infty}^{\infty} h(x)f(x)\,dx = E(R_2)$$

4. If $R \geq 0$ and $E(R) = 0$, then $R$ is *essentially* 0; that is, $P\{R = 0\} = 1$. This we can actually prove, from the previous properties. Define $R_n = 0$ if $0 \leq R < 1/n$; $R_n = 1/n$ if $R \geq 1/n$. Then $0 \leq R_n \leq R$, so that, by property 3, $E(R_n) = 0$. But $R_n$ has only two possible values, 0 and $1/n$, and so

$$E(R_n) = \sum_y y p_{R_n}(y) = 0P\{R_n = 0\} + \frac{1}{n} P\left\{R_n = \frac{1}{n}\right\}$$

Thus

$$P\left\{R_n = \frac{1}{n}\right\} = P\left\{R \geq \frac{1}{n}\right\} = 0 \qquad \text{for all } n$$

But

$$P\{R > 0\} = P\left[\bigcup_{n=1}^{\infty} \left\{R \geq \frac{1}{n}\right\}\right] \leq \sum_{n=1}^{\infty} P\left\{R \geq \frac{1}{n}\right\} = 0$$

Hence

$$P\{R = 0\} = 1$$

Notice that if $R$ is discrete, the argument is much faster: if $\sum_{x \geq 0} x p_R(x) = 0$, then $x p_R(x) = 0$ for all $x \geq 0$; hence $p_R(x) = 0$ for $x > 0$, and therefore $p_R(0) = 1$.

COROLLARY. If Var $R = 0$, then $R$ is essentially constant.

PROOF. If $m = E(R)$, then $E[(R - m)^2] = 0$, hence $P\{R = m\} = 1$.

† Since $E(R)$ is allowed to be infinite, expressions of the form $0 \cdot \infty$ will occur. The most convenient way to handle this is simply to define $0 \cdot \infty = 0$; no inconsistency will result.

5. Let $R_1, \ldots, R_n$ be *independent* random variables.
(a) If all the $R_i$ are nonnegative, then

$$E(R_1R_2 \cdots R_n) = E(R_1)E(R_2) \cdots E(R_n)$$

(b) If $E(R_i)$ is finite for all $i$ (whether or not the $R_i \geq 0$), then $E(R_1R_2 \cdots R_n)$ is finite and

$$E(R_1R_2 \cdots R_n) = E(R_1)E(R_2) \cdots E(R_n)$$

We can prove this when all the $R_i$ are discrete, if we accept certain facts about infinite series. For

$$\begin{aligned} E(R_1R_2 \cdots R_n) &= \sum_{x_1, \ldots, x_n} x_1x_2 \cdots x_n p_{12\ldots n}(x_1, \ldots, x_n) \\ &= \sum_{x_1, \ldots, x_n} x_1 \cdots x_n p_1(x_1) \cdots p_n(x_n) \end{aligned}$$

Under hypothesis (a) we may restrict the $x_i$'s to be $\geq 0$. Under hypothesis (b) the above series is absolutely convergent. Since a nonnegative or absolutely convergent series can be summed in any order, we have

$$E(R_1R_2 \cdots R_n) = \sum_{x_1} x_1p_1(x_1) \cdots \sum_{x_n} x_np_n(x_n) = E(R_1)E(R_2) \cdots E(R_n)$$

If $(R_1, \ldots, R_n)$ is absolutely continuous, the argument is similar, with sums replaced by integrals.

$$\begin{aligned} E(R_1R_2 \cdots R_n) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1 \cdots x_n f_{12\cdots n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1 \cdots x_n f_1(x_1) \cdots f_n(x_n)\, dx_1 \cdots dx_n \\ &= \int_{-\infty}^{\infty} x_1 f_1(x_1)\, dx_1 \cdots \int_{-\infty}^{\infty} x_n f_n(x_n)\, dx_n \\ &= E(R_1) \cdots E(R_n) \end{aligned}$$

6. Let $R$ be a random variable with finite mean $m$ and variance $\sigma^2$ (possibly infinite). If $a$ and $b$ are real numbers, then

$$\text{Var } (aR + b) = a^2\sigma^2$$

PROOF. Since $E(aR + b) = am + b$ by properties 1 and 2 [and (3.1.2)], we have

$$\begin{aligned} \text{Var } (aR + b) &= E[(aR + b - (am + b))^2] \\ &= E[a^2(R - m)^2] \\ &= a^2E[(R - m)^2] \qquad \text{by property 2} \\ &= a^2\sigma^2. \end{aligned}$$

7. Let $R_1, \ldots, R_n$ be independent random variables, each with finite mean. Then

$$\text{Var}\,(R_1 + \cdots + R_n) = \text{Var}\,R_1 + \cdots + \text{Var}\,R_n$$

PROOF. Let $m_i = E(R_i)$. Then

$$\text{Var}\,(R_1 + \cdots + R_n) = E\left[\left(\sum_{i=1}^{n} R_i - \sum_{i=1}^{n} m_i\right)^2\right] = E\left[\left(\sum_{i=1}^{n} (R_i - m_i)\right)^2\right]$$

If this is expanded, the "cross terms" are 0, since, if $i \neq j$,

$$\begin{aligned} E[(R_i - m_i)(R_j - m_j)] &= E(R_iR_j - m_iR_j - m_jR_i + m_im_j) \\ &= E(R_i)E(R_j) - m_iE(R_j) - m_jE(R_i) + m_im_j \\ &\qquad\qquad \text{by properties 5, 1, and 2} \\ &= 0 \qquad \text{since } E(R_i) = m_i,\ E(R_j) = m_j \end{aligned}$$

Thus

$$\text{Var}\,(R_1 + \cdots + R_n) = \sum_{i=1}^{n} E(R_i - m_i)^2 = \sum_{i=1}^{n} \text{Var}\,R_i$$

COROLLARY. If $R_1, \ldots, R_n$ are independent, each with finite mean, and $a_1, \ldots, a_n, b$ are real numbers, then

$$\text{Var}\,(a_1R_1 + \cdots + a_nR_n + b) = a_1^2\,\text{Var}\,R_1 + \cdots + a_n^2\,\text{Var}\,R_n$$

PROOF. This follows from properties 6 and 7. (Notice that $a_1R_1, \ldots, a_nR_n$ are still independent; see Problem 1.)

8. The central moments $\beta_1, \ldots, \beta_n$ $(n \geq 2)$ can be obtained from the moments $\alpha_1, \ldots, \alpha_n$, provided that $\alpha_1, \ldots, \alpha_{n-1}$ are finite and $\alpha_n$ exists.

To see this, expand $(R - m)^n$ by the binomial theorem.

$$(R - m)^n = \sum_{k=0}^{n} \binom{n}{k} R^k(-m)^{n-k}$$

Thus

$$\beta_n = E[(R - m)^n] = \sum_{k=0}^{n} \binom{n}{k} (-m)^{n-k}\alpha_k$$

Notice that since $\alpha_1, \ldots, \alpha_{n-1}$ are finite, no terms of the form $+\infty - \infty$ can appear in the summation, and thus we may take the expectation term by term, by property 1.

This result is applied most often when $n = 2$. If $R$ has finite mean [$E(R^2)$ always exists since $R^2 \geq 0$], then $(R - m)^2 = R^2 - 2mR + m^2$; hence

$$\text{Var}\,R = E(R^2) - 2mE(R) + m^2$$

That is,

$$\sigma^2 = E(R^2) - [E(R)]^2 \tag{3.3.1}$$

which is the "mean of the square" minus the "square of the mean."

9. If $E(R^k)$ is finite and $0 < j < k$, then $E(R_j)$ is also finite.

PROOF

$$\begin{aligned} |R(\omega)|^j &\leq |R(\omega)|^k \qquad \text{if } |R(\omega)| \geq 1 \\ &\leq 1 \qquad\qquad\quad \text{if } |R(\omega)| < 1 \end{aligned}$$

Thus

$$|R(\omega)|^j \leq 1 + |R(\omega)|^k \qquad \text{for all } \omega$$

Hence

$$E(|R|^j) \leq 1 + E(|R|^k) < \infty$$

and the result follows. Notice that the expectation of a random variable is finite if and only if the expectation of its absolute value is finite; see (3.1.7).

Thus in property 8, if $\alpha_{n-1}$ is finite, automatically $\alpha_1, \ldots, \alpha_{n-2}$ are finite as well.

REMARK. Properties 5 and 7 fail without the hypothesis of independence. For example, let $R_1 = R_2 = R$, where $R$ has finite mean. Then $E(R_1R_2) \neq E(R_1)E(R_2)$ since $E(R^2) - [E(R)]^2 = \text{Var } R$, which is $>0$ unless $R$ is essentially constant, by the corollary to property 4. Also, $\text{Var } (R_1 + R_2) = \text{Var } (2R) = 4 \text{ Var } R$, which is not the same as $\text{Var } R_1 + \text{Var } R_2 = 2 \text{ Var } R$ unless $R$ is essentially constant.

## PROBLEMS

1. If $R_1, \ldots, R_n$ are independent random variables, show that $a_1R_1 + b_1, \ldots, a_nR_n + b_n$ are independent for all possible choices of the constants $a_i$ and $b_i$.
2. If $R$ is normally distributed with mean $m$ and variance $\sigma^2$, evaluate the central moments of $R$ (see Problem 1, Section 3.2).
3. Let $\theta$ be uniformly distributed between 0 and $2\pi$. Define $R_1 = \cos\theta$, $R_2 = \sin\theta$. Show that $E(R_1R_2) = E(R_1)E(R_2)$, and also $\text{Var } (R_1 + R_2) = \text{Var } R_1 + \text{Var } R_2$, but $R_1$ and $R_2$ are not independent. Thus, in properties 5 and 7, the converse assertion is false.
4. If $E(R)$ exists, show that $|E(R)| \leq E(|R|)$.
5. Let $R$ be a random variable with finite mean. Indicate how and under what conditions the moments of $R$ can be obtained from the central moments. In

particular show that $E(R^2) < \infty$ if and only if Var $R < \infty$. More generally, $\alpha_n$ is finite if and only if $\beta_n$ is finite.

## 3.4 CORRELATION

If $R_1$ and $R_2$ are random variables on a given probability space, we may define *joint moments* associated with $R_1$ and $R_2$

$$\alpha_{jk} = E(R_1{}^j R_2{}^k), \qquad j, k > 0$$

and *joint central moments*

$$\beta_{jk} = E[(R_1 - m_1)^j (R_2 - m_2)^k], \qquad m_1 = E(R_1), m_2 = E(R_2)$$

We shall study $\beta_{11} = E[(R_1 - m_1)(R_2 - m_2)] = E(R_1R_2) - E(R_1)E(R_2)$, which is called the *covariance* of $R_1$ and $R_2$, written Cov $(R_1, R_2)$.

In this section we assume that $E(R_1)$ and $E(R_2)$ are finite, and $E(R_1R_2)$ exists; then the covariance of $R_1$ and $R_2$ is well defined.

***Theorem 1.*** *If $R_1$ and $R_2$ are independent, then* Cov $(R_1, R_2) = 0$, *but not conversely.*

PROOF. By property 5 of Section 3.3, independence of $R_1$ and $R_2$ implies that $E(R_1R_2) = E(R_1)E(R_2)$; hence Cov $(R_1, R_2) = 0$. An example in which Cov $(R_1, R_2) = 0$ but $R_1$ and $R_2$ are not independent is given in Problem 3 of Section 3.3.

We shall try to find out what the knowledge of the covariance of $R_1$ and $R_2$ tells us about the random variables themselves. We first establish a very useful inequality.

***Theorem 2 (Schwarz Inequality).*** *Assume that $E(R_1{}^2)$ and $E(R_2{}^2)$ are finite ($R_1$ and $R_2$ then automatically have finite mean, by property 9 of Section 3.3, and finite variance, by property 8). Then $E(R_1R_2)$ is finite, and*

$$|E(R_1R_2)|^2 \leq E(R_1{}^2)E(R_2{}^2)$$

PROOF. If $R_1$ is essentially 0, the inequality is immediate, so assume $R_1$ not essentially 0; then $E(R_1{}^2) > 0$. For any real number $x$ let

$$h(x) = E[(x\,|R_1| + |R_2|)^2] = E(R_1{}^2)x^2 + 2E(|R_1R_2|)x + E(R_2{}^2)$$

Since $h(x)$ is the expectation of a nonnegative random variable, it must be $\geq 0$ for all $x$. The quadratic equation $h(x) = 0$ has either no real roots or, at

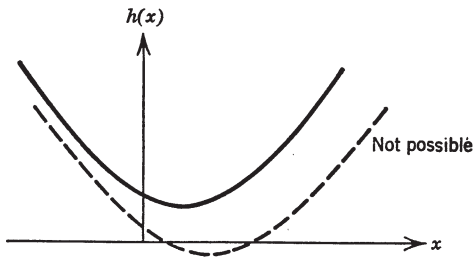


FIGURE 3.4.1 Proof of the Schwarz Inequality.

worst, one real repeated root (see Figure 3.4.1). Thus the discriminant must be $\leq 0$; hence

$$(E(|R_1R_2|))^2 \leq E(R_1^2)E(R_2^2) < \infty$$

Since $E(|R_1R_2|)$ is finite, so is $E(R_1R_2)$, by (3.1.7). Furthermore, $|E(R_1R_2)| \leq E(|R_1R_2|)$ (Problem 4, Section 3.3), and the result follows.

Now assume that $E(R_1^2)$ and $E(R_2^2)$ are finite and, in addition, that the variances $\sigma_1^2$ and $\sigma_2^2$ of $R_1$ and $R_2$ are $>0$. Define the *correlation coefficient* of $R_1$ and $R_2$ as

$$\rho(R_1, R_2) = \frac{\text{Cov}\,(R_1, R_2)}{\sigma_1\sigma_2}$$

By Theorem 1, if $R_1$ and $R_2$ are independent, they are uncorrelated; that is, $\rho(R_1, R_2) = 0$, but not conversely.

***Theorem 3.*** $-1 \leq \rho(R_1, R_2) \leq 1$.

PROOF. Apply the Schwarz inequality to $R_1 - E(R_1)$ and $R_2 - E(R_2)$.

$$|E[(R_1 - ER_1)(R_2 - ER_2)]|^2 \leq E[(R_1 - ER_1)^2]E[(R_2 - ER_2)^2]$$

Thus $|\text{Cov}\,(R_1, R_2)|^2 \leq \sigma_1^2\sigma_2^2$, and the result follows.

We shall show that $\rho$ is a *measure of linear dependence* between $R_1$ and $R_2$ [more precisely, between $R_1 - E(R_1)$ and $R_2 - E(R_2)$], in the following sense.

Let us try to estimate $R_2 - ER_2$ by a linear combination $c(R_1 - ER_1) + d$, that is, find the $c$ and $d$ that minimize

$$\begin{aligned} E\{[(R_2 - ER_2) - (c(R_1 - ER_1) + d)]^2\} \\ &= \sigma_2^2 - 2c\,\text{Cov}\,(R_1, R_2) + c^2\sigma_1^2 + d^2 \\ &= \sigma_2^2 - 2c\,\rho\sigma_1\sigma_2 + c^2\sigma_1^2 + d^2 \end{aligned}$$

Clearly we can do no better than to take $d = 0$. Now the minimum of $Ax^2 + 2Bx + D$ occurs for $x = -B/A$; hence $\sigma_1^2 c^2 - 2\rho\sigma_1\sigma_2 c + \sigma_2^2$ is minimized when

$$c = \frac{\rho\sigma_1\sigma_2}{\sigma_1^2} = \rho\frac{\sigma_2}{\sigma_1}$$

Thus the minimum expectation is $\sigma_2^2 - 2\rho^2\sigma_2^2 + \rho^2\sigma_2^2 = \sigma_2^2(1 - \rho^2)$. For a given $\sigma_2^2$, the closer $|\rho|$ is to 1, the better $R_2$ is approximated (in the mean square sense) by a linear combination $aR_1 + b$. In particular, if $|\rho| = 1$, then

$$E\left[\left(R_2 - ER_2 - \frac{\rho\sigma_2}{\sigma_1}(R_1 - ER_1)\right)^2\right] = 0$$

so that

$$R_2 - ER_2 = \frac{\rho\sigma_2}{\sigma_1}(R_1 - ER_1)$$

with probability 1.

Thus, if $|\rho| = 1$, then $R_1 - E(R_1)$ and $R_2 - E(R_2)$ are linearly dependent. (The random variables $R_1, \ldots, R_n$ are said to be linearly dependent iff there are real numbers $a_1, \ldots, a_n$, not all 0, such that $P\{a_1R_1 + \cdots + a_nR_n = 0\} = 1$.) Conversely, if $R_1 - E(R_1)$ and $R_2 - E(R_2)$ are linearly dependent, that is, if $a(R_1 - ER_1) + b(R_2 - ER_2) = 0$ with probability 1 for some constants $a$ and $b$, not both 0, then $|\rho| = 1$ (Problem 1).

## PROBLEMS

**1.** If $R_1 - E(R_1)$ and $R_2 - E(R_2)$ are linearly dependent, show that $|\rho(R_1, R_2)| = 1$.

**2.** If $aR_1 + bR_2 = c$ for some constants $a$, $b$, $c$, where $a$ and $b$ are not both 0, show that $R_1 - E(R_1)$ and $R_2 - E(R_2)$ are linearly dependent. Thus $|\rho(R_1, R_2)| = 1$ if and only if there is a line $L$ in the plane such that $(R_1(\omega), R_2(\omega))$ lies on $L$ for "almost" all $\omega$, that is, for all $\omega$ outside a set of probability 0.

**3.** Show that equality occurs in the Schwarz inequality, $|E(R_1R_2)|^2 = E(R_1^2)E(R_2^2)$, if and only if $R_1$ and $R_2$ are linearly dependent.

**4.** Prove the following results.

(a) *Schwarz inequality for sums:* For any real numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$, $(\sum_{i=1}^n a_ib_i)^2 \leq \sum_{i=1}^n a_i^2 \sum_{i=1}^n b_i^2$.

(b) *Schwarz inequality for integrals:* If $\int_a^b g^2(x)\,dx$ and $\int_a^b h^2(x)\,dx$ are finite, so is $\int_a^b g(x)h(x)\,dx$, and furthermore $(\int_a^b g(x)h(x)\,dx)^2 \leq \int_a^b g^2(x)\,dx \int_a^b h^2(x)\,dx$. HINT: show that both (a) and (b) are special cases of Theorem 2.

**5.** Show that if $R_1, \ldots, R_n$ are arbitrary random variables with $E(R_i^2)$ finite for all $i$, then

$$\text{Var}\,(R_1 + \cdots + R_n) = \sum_{i=1}^{n} \text{Var}\, R_i + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n} \text{Cov}\,(R_i, R_j)$$

## 3.5 THE METHOD OF INDICATORS

In this section we introduce a technique that in certain cases allows the expectation of a random variable to be computed quickly, without any knowledge of the distribution function. This is especially useful in situations when the distribution function is difficult to calculate.

The *indicator* of an event $A$ is a random variable $I_A$ defined as follows.

$$\begin{aligned} I_A(\omega) &= 1 \qquad \text{if } \omega \in A \\ &= 0 \qquad \text{if } \omega \notin A \end{aligned}$$

Thus $I_A = 1$ if $A$ occurs and 0 if $A$ does not occur. (Sometimes $I_A$ is called the "characteristic function" of $A$, but we do not use this terminology since we reserve the term "characteristic function" for something quite different.)

The expectation of $I_A$ is given by

$$E(I_A) = 0P\{I_A = 0\} + 1P\{I_A = 1\} = P\{I_A = 1\} = P(A)$$

The "method of indicators" simply involves expressing, if possible, a given random variable $R$ as a sum of indicators, say, $R = I_{A_1} + \cdots + I_{A_n}$. Then

$$E(R) = \sum_{j=1}^{n} E(I_{A_j}) = \sum_{j=1}^{n} P(A_j)$$

Hopefully, it will be easier to compute the $P(A_j)$ than to evaluate $E(R)$ directly.

▶ **Example 1.** Let $R$ be the number of successes in $n$ Bernoulli trials, with probability of success $p$ on a given trial; then $R$ has the binomial distribution with parameters $n$ and $p$; that is,

$$P\{R = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots . n$$

We have found by a direct evaluation that $E(R) = np$ (Problem 7, Section 3.2), but the method of indicators does the job more smoothly. Let $A_i$ be the event that there is a success on trial $i$, $i = 1, 2, \ldots, n$. Then $R = I_{A_1} + \cdots + I_{A_n}$ (note that $I_{A_i}$ may be regarded as the number of successes

on trial $i$). Thus

$$E(R) = \sum_{i=1}^{n} E(I_{A_i}) = \sum_{i=1}^{n} P(A_i) = np$$

Now, since $A_1, \ldots, A_n$ are independent, the indicators $I_{A_1}, \ldots, I_{A_n}$ are independent (Problem 1), and so there is a bonus, namely (by property 7, Section 3.3),

$$\text{Var } R = \sum_{i=1}^{n} \text{Var } I_{A_i}$$

But $I_{A_i}^2 = I_{A_i}$; hence

$$E(I_{A_i}^2) = E(I_{A_i}) = P(A_i) = p$$

Therefore

$$\begin{aligned} \text{Var } I_{A_i} &= E(I_{A_i}^2) - [E(I_{A_i})]^2 \qquad \text{by (3.3.1)} \\ &= p - p^2 = p(1-p) \end{aligned}$$

Thus

$$\text{Var } R = np(1-p) \blacktriangleleft$$

▶ **Example 2.** A single unbiased die is tossed independently $n$ times. Let $R_1$ be the number of 1's obtained, and $R_2$ the number of 2's. Find $E(R_1R_2)$.

If $A_i$ is the event that the $i$th toss results in a 1, and $B_i$ the event that the $i$th toss results in a 2, then

$$\begin{aligned} R_1 &= I_{A_1} + \cdots + I_{A_n} \\ R_2 &= I_{B_1} + \cdots + I_{B_n} \end{aligned}$$

Hence

$$E(R_1R_2) = \sum_{i,j=1}^{n} E(I_{A_i}I_{B_j})$$

Now if $i \neq j$, $I_{A_i}$ and $I_{B_j}$ are independent (see Problem 1); hence

$$E(I_{A_i}I_{B_j}) = E(I_{A_i})E(I_{B_j}) = P(A_i)P(B_j) = \tfrac{1}{36}$$

If $i = j$, $A_i$ and $B_i$ are disjoint, since the $i$th toss cannot simultaneously result in a 1 and a 2. Thus $I_{A_i}I_{B_i} = I_{A_i \cap B_i} = 0$ (see Problem 2). Thus

$$E(R_1R_2) = \frac{n(n-1)}{36}$$

since there are $n(n-1)$ ordered pairs $(i,j)$ of integers $\in\{1, 2, \ldots, n\}$ such that $i \neq j$.

Note that the $I_{A_i}I_{B_j}$, $i, j = 1, \ldots, n$, are not independent [for instance, if $I_{A_1}(\omega)I_{B_2}(\omega) = 1$, then $I_{A_2}(\omega)I_{B_3}(\omega)$ must be 0], so that we cannot compute the variance of $R_1R_2$ in the same way as in Example 1. ◀

## PROBLEMS

1. If the events $A_1, \ldots, A_n$ are independent, show that the indicators $I_{A_1}, \ldots, I_{A_n}$ are independent random variables, and conversely.
2. Establish the following properties of indicators:
   (a) $I_\Omega = 1, \quad I_\varnothing = 0$
   (b) $I_{A\cap B} = I_A I_B, \quad I_{A\cup B} = I_A + I_B - I_{A\cap B}$
   (c) $I_{\bigcup_{i=1}^{\infty} A_i} = \sum_{i=1}^{\infty} I_{A_i}$ if the $A_i$ are disjoint
   (d) If $A_1, A_2, \ldots$ is an expanding sequence of events ($A_n \subset A_{n+1}$ for all $n$) and $\bigcup_{n=1}^{\infty} A_n = A$, or if $A_1, A_2, \ldots$ is a contracting sequence ($A_{n+1} \subset A_n$ for all $n$) and $\bigcap_{n=1}^{\infty} A_n = A$, then $I_{A_n} \to I_A$; that is, $\lim_{n\to\infty} I_{A_n}(\omega) = I_A(\omega)$ for all $\omega$.
3. In Example 2, find the joint probability function of $R_1$ and $R_2$. Notice how unwieldy is the direct expression for $E(R_1R_2)$.

$$E(R_1R_2) = \sum_{j,k=0}^{n} jkP\{R_1 = j, R_2 = k\}$$

4. In a sequence of $n$ Bernoulli trials, let $R_0$ be the number of times a success is followed immediately by a failure. For example, if $n = 7$ and $\omega = (S\overline{SF}F\overline{SF}S)$, then $R_0(\omega) = 2$, as indicated. Find $E(R_0)$.
5. Find Var $R_0$ in Problem 4.
6. 100 balls are tossed independently and at random into 50 boxes. Let $R$ be the number of empty boxes. Find $E(R)$.

## 3.6 SOME PROPERTIES OF THE NORMAL DISTRIBUTION

Let $R_1$ be normally distributed with mean $m$ and variance $\sigma^2$.

$$f_1(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m)^2/2\sigma^2}$$

If $R_2 = aR_1 + b$, $a \neq 0$, we shall show that $R_2$ is also normally distributed [necessarily $E(R_2) = am + b$, Var $R_2 = a^2\sigma^2$ by properties 1, 2, and 6 of Section 3.3].

We may use the technique of Section 2.4 to find the density of $R_2$. $R_2 = y$ corresponds to $R_1 = h(y) = (y - b)/a$. Thus

$$f_2(y) = f_1(h(y))\,|h'(y)| = \frac{1}{|a|} f_1\left(\frac{y-b}{a}\right)$$

$$= \frac{1}{\sqrt{2\pi}\,|a|\,\sigma} \exp\left[-\frac{(y-(am+b))^2}{2a^2\sigma^2}\right]$$

so that $R_2$ has the normal density with mean $am + b$ and variance $a^2\sigma^2$. We may use this result in the calculation of probabilities of events involving a normally distributed random variable. If $R$ has the normal density with $E(R) = m$, Var $R = \sigma^2$, then

$$P\{a \leq R \leq b\} = \int_a^b \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m)^2/2\sigma^2}\,dx$$

One must resort to tables to evaluate this. The point we wish to bring out is that, regardless of $m$ and $\sigma^2$, only one table is needed, namely, that of the normal distribution function when $m = 0$, $\sigma^2 = 1$; that is,

$$F^*(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt$$

For if $R$ is normally distributed with $E(R) = m$, Var $R = \sigma^2$, then $R^* = (R - m)/\sigma$ is normally distributed with $E(R^*) = 0$ and Var $R^* = 1$. Thus

$$\begin{aligned} P\{a \leq R \leq b\} &= P\left\{\frac{a-m}{\sigma} \leq R^* \leq \frac{b-m}{\sigma}\right\} \\ &= F^*\left(\frac{b-m}{\sigma}\right) - F^*\left(\frac{a-m}{\sigma}\right) \end{aligned}$$

A brief table of values of $F^*$ is given at the end of the book.

REMARK. If a random variable has a density function $f$ that is symmetrical about 0 [i.e., an even function: $f(-x) = f(x)$], then the distribution function has the property that $F(-x) = 1 - F(x)$. For (see Figure 3.6.1)

$$\begin{aligned} F(-x) = P\{R \leq -x\} &= \int_{-\infty}^{-x} f(t)\,dt = \int_x^{\infty} f(t)\,dt \\ &= P\{R > x\} = 1 - F(x) \end{aligned}$$

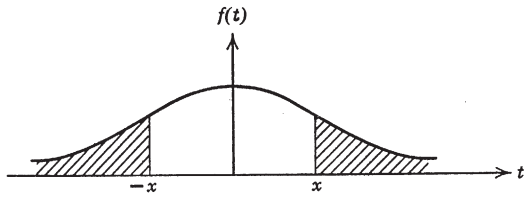


FIGURE 3.6.1 Symmetrical Density.

In particular, the distribution function $F^*$ has this property, and thus once the values of $F^*(x)$ for positive $x$ are known, the values of $F^*(x)$ for negative $x$ are determined.

## PROBLEMS

**1.** Let $R$ be normally distributed with $m = 1$, $\sigma^2 = 9$.
(a) Find $P\{-.5 \leq R \leq 4\}$
(b) If $P\{R \geq c\} = .9$, find $c$.

**2.** If $R$ is normally distributed and $k$ is a positive real number, show that $P\{|R - m| \geq k\sigma\}$ does not depend on $m$ or $\sigma$; thus one can speak unambiguously of the "probability that a normally distributed random variable lies at least $k$ standard deviations from its mean." Show that when $k = 1.96$, the probability is .05.

## 3.7 CHEBYSHEV'S INEQUALITY AND THE WEAK LAW OF LARGE NUMBERS

In this section we are going to prove a result that corresponds to the physical statement that the arithmetic average of a very large number of independent observations of a random variable $R$ is very likely to be very close to $E(R)$. We first establish a quantitative result about the variance as a measure of dispersion.

***Theorem 1.***

(a) *Let $R$ be a nonnegative random variable, and $b$ a positive real number. Then*

$$P\{R \geq b\} \leq \frac{E(R)}{b}$$

PROOF. We first consider the absolutely continuous case. We have

$$E(R) = \int_{-\infty}^{\infty} x f_R(x)\, dx = \int_0^{\infty} x f_R(x)\, dx$$

since $R \geq 0$, so that $f_R(x) = 0$, $x < 0$. Now if we drop the integral from 0 to $b$, we get something smaller.

$$E(R) \geq \int_b^{\infty} x f_R(x)\, dx$$

Since $x \geq b$,

$$\int_b^\infty x f_R(x)\,dx \geq \int_b^\infty b f_R(x)\,dx = bP\{R \geq b\}$$

This is the desired result.

The general proof is based on the same idea. Let $A_b = \{R \geq b\}$; then $R \geq RI_{A_b}$. For if $\omega \notin A_b$, this says simply that $R(\omega) \geq 0$; if $\omega \in A_b$, it says that $R(\omega) \geq R(\omega)$. Thus $E(R) \geq E(RI_{A_b})$. But $RI_{A_b} \geq bI_{A_b}$, since $\omega \in A_b$ implies that $R(\omega) \geq b$. Thus

$$E(R) \geq E(RI_{A_b}) \geq E(bI_{A_b}) = bE(I_{A_b}) = bP(A_b)$$

Consequently $P(A_b) \leq E(R)/b$, as desired.

(b) *Let $R$ be an arbitrary random variable, $c$ any real number, and $\varepsilon$ and $m$ positive real numbers. Then*

$$P\{|R - c| \geq \varepsilon\} \leq \frac{E[|R - c|^m]}{\varepsilon^m}$$

PROOF.

$$P\{|R - c| \geq \varepsilon\} = P\{|R - c|^m \geq \varepsilon^m\} \leq \frac{E[|R - c|^m]}{\varepsilon^m} \qquad \text{by (a)}$$

(c) *If $R$ has finite mean $m$ and finite variance $\sigma^2 > 0$, and $k$ is a positive real number, then*

$$P\{|R - m| \geq k\sigma\} \leq \frac{1}{k^2}$$

PROOF. This follows from (b) with $c = m$, $\varepsilon = k\sigma$, $m = 2$.

All three parts of Theorem 1 go under the name of *Chebyshev's inequality*. Part (c) says that the probability that a random variable will fall $k$ or more standard deviations from its mean is $\leq 1/k^2$. Notice that nothing at all is said about the distribution function of $R$; Chebyshev's inequality is therefore quite a general statement. When applied to a particular case, however, it may be quite weak. For example, let $R$ be normally distributed with mean $m$ and variance $\sigma^2$. Then (Problem 2, Section 3.6) $P\{|R - m| \geq 1.96\sigma\} = .05$. In this case Chebyshev's inequality says only that

$$P\{|R - m| \geq 1.96\sigma\} \leq \frac{1}{(1.96)^2} = .26$$

which is a much weaker statement. The strength of Chebyshev's inequality lies in its universality.

We are now ready for the main result.

***Theorem 2.*** **(*Weak Law of Large Numbers*).** *For each $n = 1, 2, \ldots,$ suppose that $R_1, R_2, \ldots, R_n$ are independent random variables on a given probability space, each having finite mean and variance. Assume that the variances are uniformly bounded; that is, assume that there is some finite positive number $M$ such that $\sigma_i^2 \leq M$ for all $i$. Let $S_n = \sum_{i=1}^{n} R_i$. Then, for any $\varepsilon > 0$,*

$$P\left\{\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right\} \to 0 \qquad \text{as } n \to \infty$$

Before proving the theorem, we consider two cases of interest.

SPECIAL CASES

1. Suppose that $E(R_i) = m$ for all $i$, and Var $R_i = \sigma^2$ for all $i$. Then

$$E(S_n) = \sum_{i=1}^{n} E(R_i) = nm$$

$$\frac{S_n - E(S_n)}{n} = \frac{S_n}{n} - m = \frac{R_1 + \cdots + R_n}{n} - m$$

Therefore, for any arbitrary $\varepsilon > 0$, there is for large $n$ a high probability that the arithmetic average and the expectation $m$ will differ by $< \varepsilon$.

This case covers the situation when $R_1, R_2, \ldots, R_n$ are independent observations of a given random variable $R$. All this means is that $R_1, \ldots, R_n$ are independent, and the $R_i$ all have the same distribution function, namely, $F_R$. In particular, $E(R_i) = E(R)$, so that for large $n$ there is a high probability that $(R_1 + \cdots + R_n)/n$ and $E(R)$ will differ by $< \varepsilon$.

2. Consider a sequence of Bernoulli trials, and let $R_i$ be the number of successes on trial $i$; that is, $R_i = I_{A_i}$, where $A_i = \{\text{success on trial } i\}$. Then $(R_1 + \cdots + R_n)/n$ is the relative frequency of successes in $n$ trials. Now $E(R_i) = P(A_i) = p$, the probability of success on a given trial, so that for large $n$ there is a high probability that the relative frequency will differ from $p$ by $< \varepsilon$.

PROOF OF THEOREM 2. By the second form of Chebyshev's inequality [part (b) of Theorem 1], with $R = (S_n - E(S_n))/n$, $c = 0$, and $m = 2$, we have

$$P\left\{\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2} E\left[\left(\frac{S_n - E(S_n)}{n}\right)^2\right] = \frac{1}{n^2\varepsilon^2} \text{Var } S_n$$

But since $R_1, \ldots, R_n$ are independent,

$$\text{Var } S_n = \sum_{i=1}^{n} \text{Var } R_i \qquad \text{by property 7 of Section 3.3}$$

$$\leq nM \qquad \text{by hypothesis}$$

Thus

$$P\left\{\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right\} \leq \frac{nM}{n^2\varepsilon^2} = \frac{M}{n\varepsilon} \to 0$$

REMARK. If a coin with probability $p$ of heads is tossed indefinitely, the successive tosses being independent, we expect that as a practical matter the relative frequency of heads will converge, in the ordinary sense of convergence of a sequence of real numbers, to $p$. This is a somewhat stronger statement than the weak law of large numbers, which says that for large $n$ the relative frequency of heads in $n$ trials is very likely to be very close to $p$. The first statement, when properly formulated, becomes the *strong law of large numbers*, which we shall examine in detail later.

## PROBLEMS

**1.** Let $R$ have the exponential density $f(x) = e^{-x}, x \geq 0; f(x) = 0, x < 0$. Evaluate $P\{|R - m| \geq k\sigma\}$ and compare with the Chebyshev bound.

**2.** Suppose that we have a sequence of random variables $R_n$ such that $P\{R_n = e^n\} = 1/n, P\{R_n = 0\} = 1 - 1/n, n = 1, 2, \ldots$.

(a) State and prove a theorem that expresses the fact that for large $n$, $R_n$ is very likely to be 0.

(b) Show that $E(R_n{}^k) \to \infty$ as $n \to \infty$ for any $k > 0$.

**3.** Suppose that $R_n$ is the amount you win on trial $n$ in a game of chance. Assume that the $R_i$ are independent random variables, each with finite mean $m$ and finite variance $\sigma^2$. Make the realistic assumption that $m < 0$. Show that $P\{(R_1 + \cdots + R_n)/n < m/2\} \to 1$ as $n \to \infty$. What is the moral of this result?