Chapter 6

Expected Value and Variance

6.1 Expected Value of Discrete Random Variables

When a large collection of numbers is assembled, as in a census, we are usually interested not in the individual numbers, but rather in certain descriptive quantities such as the average or the median. In general, the same is true for the probability distribution of a numerically-valued random variable. In this and in the next section, we shall discuss two such descriptive quantities: the *expected value* and the *variance.* Both of these quantities apply only to numerically-valued random variables, and so we assume, in these sections, that all random variables have numerical values. To give some intuitive justification for our definition, we consider the following game.

Average Value

A die is rolled. If an odd number turns up, we win an amount equal to this number; if an even number turns up, we lose an amount equal to this number. For example, if a two turns up we lose 2, and if a three comes up we win 3. We want to decide if this is a reasonable game to play. We first try simulation. The program **Die** carries out this simulation.

The program prints the frequency and the relative frequency with which each outcome occurs. It also calculates the average winnings. We have run the program twice. The results are shown in Table 6.1.

In the first run we have played the game 100 times. In this run our average gain is $-.57$. It looks as if the game is unfavorable, and we wonder how unfavorable it really is. To get a better idea, we have played the game 10,000 times. In this case our average gain is $-.4949$.

We note that the relative frequency of each of the six possible outcomes is quite close to the probability $1/6$ for this outcome. This corresponds to our frequency interpretation of probability. It also suggests that for very large numbers of plays, our average gain should be

$$\mu \quad = \quad 1\Big(\frac{1}{6}\Big) - 2\Big(\frac{1}{6}\Big) + 3\Big(\frac{1}{6}\Big) - 4\Big(\frac{1}{6}\Big) + 5\Big(\frac{1}{6}\Big) - 6\Big(\frac{1}{6}\Big)$$

Winning	n = 100		n = 10000	
	Frequency	Relative Frequency	Frequency	Relative Frequency
1	17	.17	1681	.1681
-2	17	.17	1678	.1678
3	16	.16	1626	.1626
-4	18	.18	1696	.1696
5	16	.16	1686	.1686
-6	16	.16	1633	.1633

Table 6.1: Frequencies for dice game.

$$= \frac{9}{6} - \frac{12}{6} = -\frac{3}{6} = -.5 \ .$$

This agrees quite well with our average gain for 10,000 plays.

We note that the value we have chosen for the average gain is obtained by taking the possible outcomes, multiplying by the probability, and adding the results. This suggests the following definition for the expected outcome of an experiment.

Expected Value

Definition 6.1 Let X be a numerically-valued discrete random variable with sample space Ω and distribution function $m(x)$. The *expected value* $E(X)$ is defined by

$$E(X) = \sum_{x \in \Omega} x m(x) \ ,$$

provided this sum converges absolutely. We often refer to the expected value as the *mean,* and denote $E(X)$ by μ for short. If the above sum does not converge absolutely, then we say that X does not have an expected value. □

Example 6.1 Let an experiment consist of tossing a fair coin three times. Let X denote the number of heads which appear. Then the possible values of X are $0, 1, 2$ and 3. The corresponding probabilities are $1/8, 3/8, 3/8$, and $1/8$. Thus, the expected value of X equals

$$0\left(\frac{1}{8}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) = \frac{3}{2} \ .$$

Later in this section we shall see a quicker way to compute this expected value, based on the fact that X can be written as a sum of simpler random variables. □

Example 6.2 Suppose that we toss a fair coin until a head first comes up, and let X represent the number of tosses which were made. Then the possible values of X are $1, 2, \ldots$, and the distribution function of X is defined by

$$m(i) = \frac{1}{2^i} \ .$$

(This is just the geometric distribution with parameter 1/2.) Thus, we have

$$
\begin{aligned}
E(X) &= \sum_{i=1}^{\infty} i \frac{1}{2^i} \\
&= \sum_{i=1}^{\infty} \frac{1}{2^i} + \sum_{i=2}^{\infty} \frac{1}{2^i} + \cdots \\
&= 1 + \frac{1}{2} + \frac{1}{2^2} + \cdots \\
&= 2\ .
\end{aligned}
$$

□

Example 6.3 (Example 6.2 continued) Suppose that we flip a coin until a head first appears, and if the number of tosses equals n, then we are paid 2^n dollars. What is the expected value of the payment?

We let Y represent the payment. Then,

$$P(Y = 2^n) = \frac{1}{2^n}\ ,$$

for $n \geq 1$. Thus,

$$E(Y) = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n}\ ,$$

which is a divergent sum. Thus, Y has no expectation. This example is called the *St. Petersburg Paradox*. The fact that the above sum is infinite suggests that a player should be willing to pay any fixed amount per game for the privilege of playing this game. The reader is asked to consider how much he or she would be willing to pay for this privilege. It is unlikely that the reader's answer is more than 10 dollars; therein lies the paradox.

In the early history of probability, various mathematicians gave ways to resolve this paradox. One idea (due to G. Cramer) consists of assuming that the amount of money in the world is finite. He thus assumes that there is some fixed value of n such that if the number of tosses equals or exceeds n, the payment is 2^n dollars. The reader is asked to show in Exercise 20 that the expected value of the payment is now finite.

Daniel Bernoulli and Cramer also considered another way to assign value to the payment. Their idea was that the value of a payment is some function of the payment; such a function is now called a utility function. Examples of reasonable utility functions might include the square-root function or the logarithm function. In both cases, the value of $2n$ dollars is less than twice the value of n dollars. It can easily be shown that in both cases, the expected utility of the payment is finite (see Exercise 20). □

Example 6.4 Let T be the time for the first success in a Bernoulli trials process. Then we take as sample space Ω the integers $1,\ 2,\ \ldots$ and assign the geometric distribution

$$m(j) = P(T = j) = q^{j-1}p\ .$$

Thus,

$$\begin{aligned} E(T) &= 1 \cdot p + 2qp + 3q^2p + \cdots \\ &= p(1 + 2q + 3q^2 + \cdots)\ . \end{aligned}$$

Now if $|x| < 1$, then

$$1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}\ .$$

Differentiating this formula, we get

$$1 + 2x + 3x^2 + \cdots = \frac{1}{(1-x)^2}\ ,$$

so

$$E(T) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}\ .$$

In particular, we see that if we toss a fair coin a sequence of times, the expected time until the first heads is $1/(1/2) = 2$. If we roll a die a sequence of times, the expected number of rolls until the first six is $1/(1/6) = 6$. □

Interpretation of Expected Value

In statistics, one is frequently concerned with the average value of a set of data. The following example shows that the ideas of average value and expected value are very closely related.

Example 6.5 The heights, in inches, of the women on the Swarthmore basketball team are 5' 9", 5' 9", 5' 6", 5' 8", 5' 11", 5' 5", 5' 7", 5' 6", 5' 6", 5' 7", 5' 10", and 6' 0".

A statistician would compute the average height (in inches) as follows:

$$\frac{69 + 69 + 66 + 68 + 71 + 65 + 67 + 66 + 66 + 67 + 70 + 72}{12} = 67.9\ .$$

One can also interpret this number as the expected value of a random variable. To see this, let an experiment consist of choosing one of the women at random, and let X denote her height. Then the expected value of X equals 67.9. □

Of course, just as with the frequency interpretation of probability, to interpret expected value as an average outcome requires further justification. We know that for any finite experiment the average of the outcomes is not predictable. However, we shall eventually prove that the average will usually be close to $E(X)$ if we repeat the experiment a large number of times. We first need to develop some properties of the expected value. Using these properties, and those of the concept of the variance

X	Y
HHH	1
HHT	2
HTH	3
HTT	2
THH	2
THT	3
TTH	2
TTT	1

Table 6.2: Tossing a coin three times.

to be introduced in the next section, we shall be able to prove the *Law of Large Numbers.* This theorem will justify mathematically both our frequency concept of probability and the interpretation of expected value as the average value to be expected in a large number of experiments.

Expectation of a Function of a Random Variable

Suppose that X is a discrete random variable with sample space Ω, and $\phi(x)$ is a real-valued function with domain Ω. Then $\phi(X)$ is a real-valued random variable. One way to determine the expected value of $\phi(X)$ is to first determine the distribution function of this random variable, and then use the definition of expectation. However, there is a better way to compute the expected value of $\phi(X)$, as demonstrated in the next example.

Example 6.6 Suppose a coin is tossed 9 times, with the result

$$HHHTTTTHT .$$

The first set of three heads is called a *run*. There are three more runs in this sequence, namely the next four tails, the next head, and the next tail. We do not consider the first two tosses to constitute a run, since the third toss has the same value as the first two.

Now suppose an experiment consists of tossing a fair coin three times. Find the expected number of runs. It will be helpful to think of two random variables, X and Y, associated with this experiment. We let X denote the sequence of heads and tails that results when the experiment is performed, and Y denote the number of runs in the outcome X. The possible outcomes of X and the corresponding values of Y are shown in Table 6.2.

To calculate $E(Y)$ using the definition of expectation, we first must find the distribution function $m(y)$ of Y i.e., we group together those values of X with a common value of Y and add their probabilities. In this case, we calculate that the distribution function of Y is: $m(1) = 1/4$, $m(2) = 1/2$, and $m(3) = 1/4$. One easily finds that $E(Y) = 2$.

Now suppose we didn't group the values of X with a common Y-value, but instead, for each X-value x, we multiply the probability of x and the corresponding value of Y, and add the results. We obtain

$$1\left(\frac{1}{8}\right)+2\left(\frac{1}{8}\right)+3\left(\frac{1}{8}\right)+2\left(\frac{1}{8}\right)+2\left(\frac{1}{8}\right)+3\left(\frac{1}{8}\right)+2\left(\frac{1}{8}\right)+1\left(\frac{1}{8}\right) ,$$

which equals 2.

This illustrates the following general principle. If X and Y are two random variables, and Y can be written as a function of X, then one can compute the expected value of Y using the distribution function of X. □

Theorem 6.1 If X is a discrete random variable with sample space Ω and distribution function $m(x)$, and if $\phi : \Omega \to \mathrm{R}$ is a function, then

$$E(\phi(X)) = \sum_{x\in\Omega} \phi(x)m(x) ,$$

provided the series converges absolutely. □

The proof of this theorem is straightforward, involving nothing more than grouping values of X with a common Y-value, as in Example 6.6.

The Sum of Two Random Variables

Many important results in probability theory concern sums of random variables. We first consider what it means to add two random variables.

Example 6.7 We flip a coin and let X have the value 1 if the coin comes up heads and 0 if the coin comes up tails. Then, we roll a die and let Y denote the face that comes up. What does $X + Y$ mean, and what is its distribution? This question is easily answered in this case, by considering, as we did in Chapter 4, the joint random variable $Z = (X, Y)$, whose outcomes are ordered pairs of the form (x, y), where $0 \le x \le 1$ and $1 \le y \le 6$. The description of the experiment makes it reasonable to assume that X and Y are independent, so the distribution function of Z is uniform, with $1/12$ assigned to each outcome. Now it is an easy matter to find the set of outcomes of $X + Y$, and its distribution function. □

In Example 6.1, the random variable X denoted the number of heads which occur when a fair coin is tossed three times. It is natural to think of X as the sum of the random variables X_1, X_2, X_3, where X_i is defined to be 1 if the ith toss comes up heads, and 0 if the ith toss comes up tails. The expected values of the X_i's are extremely easy to compute. It turns out that the expected value of X can be obtained by simply adding the expected values of the X_i's. This fact is stated in the following theorem.

Theorem 6.2 Let X and Y be random variables with finite expected values. Then

$$E(X+Y) = E(X) + E(Y)\ ,$$

and if c is any constant, then

$$E(cX) = cE(X)\ .$$

Proof. Let the sample spaces of X and Y be denoted by Ω_X and Ω_Y, and suppose that

$$\Omega_X = \{x_1, x_2, \ldots\}$$

and

$$\Omega_Y = \{y_1, y_2, \ldots\}\ .$$

Then we can consider the random variable $X+Y$ to be the result of applying the function $\phi(x,y) = x+y$ to the joint random variable (X,Y). Then, by Theorem 6.1, we have

$$\begin{aligned}
E(X+Y) &= \sum_j \sum_k (x_j + y_k) P(X = x_j,\ Y = y_k) \\
&= \sum_j \sum_k x_j P(X = x_j,\ Y = y_k) + \sum_j \sum_k y_k P(X = x_j,\ Y = y_k) \\
&= \sum_j x_j P(X = x_j) + \sum_k y_k P(Y = y_k)\ .
\end{aligned}$$

The last equality follows from the fact that

$$\sum_k P(X = x_j,\ Y = y_k) = P(X = x_j)$$

and

$$\sum_j P(X = x_j,\ Y = y_k) = P(Y = y_k)\ .$$

Thus,

$$E(X+Y) = E(X) + E(Y)\ .$$

If c is any constant,

$$\begin{aligned}
E(cX) &= \sum_j c x_j P(X = x_j) \\
&= c \sum_j x_j P(X = x_j) \\
&= cE(X)\ .
\end{aligned}$$

$\square$

X			Y
a	b	c	3
a	c	b	1
b	a	c	1
b	c	a	0
c	a	b	0
c	b	a	1

Table 6.3: Number of fixed points.

It is easy to prove by mathematical induction that *the expected value of the sum of any finite number of random variables is the sum of the expected values of the individual random variables.*

It is important to note that mutual independence of the summands was not needed as a hypothesis in the Theorem 6.2 and its generalization. The fact that expectations add, whether or not the summands are mutually independent, is sometimes referred to as the First Fundamental Mystery of Probability.

Example 6.8 Let Y be the number of fixed points in a random permutation of the set $\{a, b, c\}$. To find the expected value of Y, it is helpful to consider the basic random variable associated with this experiment, namely the random variable X which represents the random permutation. There are six possible outcomes of X, and we assign to each of them the probability 1/6 see Table 6.3. Then we can calculate $E(Y)$ using Theorem 6.1, as

$$3\Bigl(\frac{1}{6}\Bigr) + 1\Bigl(\frac{1}{6}\Bigr) + 1\Bigl(\frac{1}{6}\Bigr) + 0\Bigl(\frac{1}{6}\Bigr) + 0\Bigl(\frac{1}{6}\Bigr) + 1\Bigl(\frac{1}{6}\Bigr) = 1\ .$$

We now give a very quick way to calculate the average number of fixed points in a random permutation of the set $\{1, 2, 3, \ldots, n\}$. Let Z denote the random permutation. For each i, $1 \le i \le n$, let X_i equal 1 if Z fixes i, and 0 otherwise. So if we let F denote the number of fixed points in Z, then

$$F = X_1 + X_2 + \cdots + X_n\ .$$

Therefore, Theorem 6.2 implies that

$$E(F) = E(X_1) + E(X_2) + \cdots + E(X_n)\ .$$

But it is easy to see that for each i,

$$E(X_i) = \frac{1}{n}\ ,$$

so

$$E(F) = 1\ .$$

This method of calculation of the expected value is frequently very useful. It applies whenever the random variable in question can be written as a sum of simpler random variables. We emphasize again that it is not necessary that the summands be mutually independent. □

Bernoulli Trials

Theorem 6.3 Let S_n be the number of successes in n Bernoulli trials with probability p for success on each trial. Then the expected number of successes is np. That is,

$$E(S_n) = np\ .$$

Proof. Let X_j be a random variable which has the value 1 if the jth outcome is a success and 0 if it is a failure. Then, for each X_j,

$$E(X_j) = 0 \cdot (1 - p) + 1 \cdot p = p\ .$$

Since

$$S_n = X_1 + X_2 + \cdots + X_n\ ,$$

and the expected value of the sum is the sum of the expected values, we have

$$\begin{aligned} E(S_n) &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= np\ . \end{aligned}$$

□

Poisson Distribution

Recall that the Poisson distribution with parameter λ was obtained as a limit of binomial distributions with parameters n and p, where it was assumed that $np = \lambda$, and $n \to \infty$. Since for each n, the corresponding binomial distribution has expected value λ, it is reasonable to guess that the expected value of a Poisson distribution with parameter λ also has expectation equal to λ. This is in fact the case, and the reader is invited to show this (see Exercise 21).

Independence

If X and Y are two random variables, it is not true in general that $E(X \cdot Y) = E(X)E(Y)$. However, this is true if X and Y are *independent.*

Theorem 6.4 If X and Y are independent random variables, then

$$E(X \cdot Y) = E(X)E(Y)\ .$$

Proof. Suppose that

$$\Omega_X = \{x_1, x_2, \ldots\}$$

and

$$\Omega_Y = \{y_1, y_2, \ldots\}$$

are the sample spaces of X and Y, respectively. Using Theorem 6.1, we have

$$E(X \cdot Y) = \sum_j \sum_k x_j y_k P(X = x_j,\ Y = y_k)\ .$$

But if X and Y are independent,

$$P(X = x_j, Y = y_k) = P(X = x_j)P(Y = y_k)\ .$$

Thus,

$$\begin{aligned} E(X \cdot Y) &= \sum_j \sum_k x_j y_k P(X = x_j)P(Y = y_k) \\ &= \left(\sum_j x_j P(X = x_j)\right)\left(\sum_k y_k P(Y = y_k)\right) \\ &= E(X)E(Y)\ . \end{aligned}$$

□

Example 6.9 A coin is tossed twice. $X_i = 1$ if the ith toss is heads and 0 otherwise. We know that X_1 and X_2 are independent. They each have expected value 1/2. Thus $E(X_1 \cdot X_2) = E(X_1)E(X_2) = (1/2)(1/2) = 1/4$. □

We next give a simple example to show that the expected values need not multiply if the random variables are not independent.

Example 6.10 Consider a single toss of a coin. We define the random variable X to be 1 if heads turns up and 0 if tails turns up, and we set $Y = 1 - X$. Then $E(X) = E(Y) = 1/2$. But $X \cdot Y = 0$ for either outcome. Hence, $E(X \cdot Y) = 0 \neq E(X)E(Y)$. □

We return to our records example of Section 3.1 for another application of the result that the expected value of the sum of random variables is the sum of the expected values of the individual random variables.

Records

Example 6.11 We start keeping snowfall records this year and want to find the expected number of records that will occur in the next n years. The first year is necessarily a record. The second year will be a record if the snowfall in the second year is greater than that in the first year. By symmetry, this probability is 1/2. More generally, let X_j be 1 if the jth year is a record and 0 otherwise. To find $E(X_j)$, we need only find the probability that the jth year is a record. But the record snowfall for the first j years is equally likely to fall in any one of these years, so $E(X_j) = 1/j$. Therefore, if S_n is the total number of records observed in the first n years,

$$E(S_n) = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\ .$$

This is the famous *divergent harmonic series.* It is easy to show that

$$E(S_n) \sim \log n$$

as $n \to \infty$. A more accurate approximation to $E(S_n)$ is given by the expression

$$\log n + \gamma + \frac{1}{2n}\ ,$$

where γ denotes Euler's constant, and is approximately equal to .5772.

Therefore, in ten years the expected number of records is approximately 2.9298; the exact value is the sum of the first ten terms of the harmonic series which is 2.9290. □

Craps

Example 6.12 In the game of craps, the player makes a bet and rolls a pair of dice. If the sum of the numbers is 7 or 11 the player wins, if it is 2, 3, or 12 the player loses. If any other number results, say r, then r becomes the player's point and he continues to roll until either r or 7 occurs. If r comes up first he wins, and if 7 comes up first he loses. The program **Craps** simulates playing this game a number of times.

We have run the program for 1000 plays in which the player bets 1 dollar each time. The player's average winnings were $-.006$. The game of craps would seem to be only slightly unfavorable. Let us calculate the expected winnings on a single play and see if this is the case. We construct a two-stage tree measure as shown in Figure 6.1.

The first stage represents the possible sums for his first roll. The second stage represents the possible outcomes for the game if it has not ended on the first roll. In this stage we are representing the possible outcomes of a sequence of rolls required to determine the final outcome. The branch probabilities for the first stage are computed in the usual way assuming all 36 possibilites for outcomes for the pair of dice are equally likely. For the second stage we assume that the game will eventually end, and we compute the conditional probabilities for obtaining either the point or a 7. For example, assume that the player's point is 6. Then the game will end when one of the eleven pairs, $(1,5)$, $(2,4)$, $(3,3)$, $(4,2)$, $(5,1)$, $(1,6)$, $(2,5)$, $(3,4)$, $(4,3)$, $(5,2)$, $(6,1)$, occurs. We assume that each of these possible pairs has the same probability. Then the player wins in the first five cases and loses in the last six. Thus the probability of winning is $5/11$ and the probability of losing is $6/11$. From the path probabilities, we can find the probability that the player wins 1 dollar; it is $244/495$. The probability of losing is then $251/495$. Thus if X is his winning for

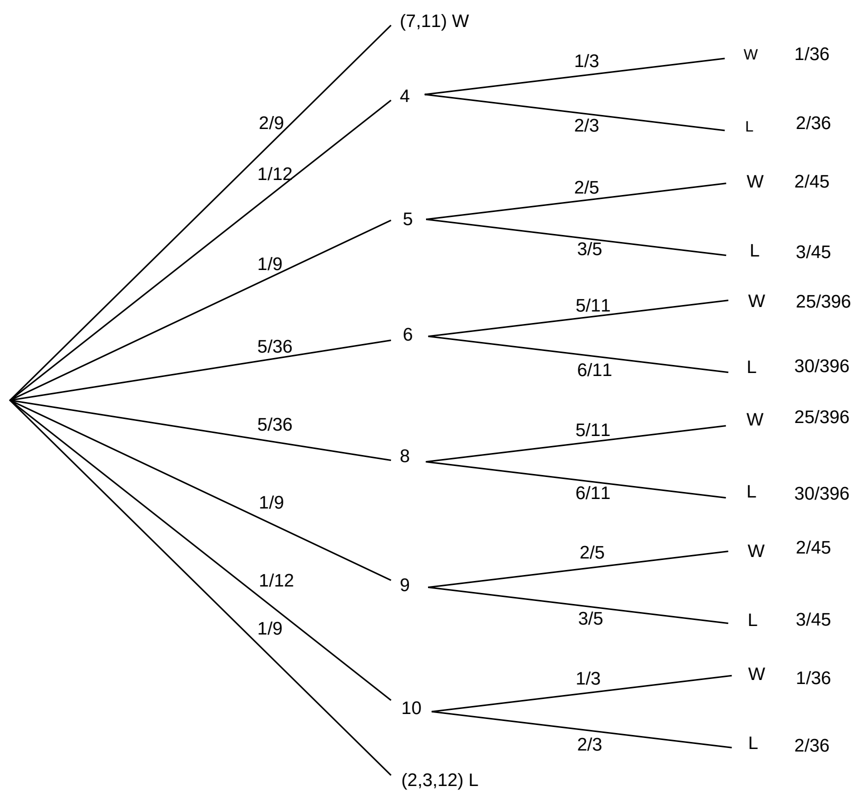

Figure 6.1: Tree measure for craps.

a dollar bet,

$$\begin{aligned} E(X) &= 1\Big(\frac{244}{495}\Big) + (-1)\Big(\frac{251}{495}\Big) \\ &= -\frac{7}{495} \approx -.0141 . \end{aligned}$$

The game is unfavorable, but only slightly. The player's expected gain in n plays is $-n(.0141)$. If n is not large, this is a small expected loss for the player. The casino makes a large number of plays and so can afford a small average gain per play and still expect a large profit. □

Roulette

Example 6.13 In Las Vegas, a roulette wheel has 38 slots numbered 0, 00, 1, 2, . . . , 36. The 0 and 00 slots are green, and half of the remaining 36 slots are red and half are black. A croupier spins the wheel and throws an ivory ball. If you bet 1 dollar on red, you win 1 dollar if the ball stops in a red slot, and otherwise you lose a dollar. We wish to calculate the expected value of your winnings, if you bet 1 dollar on red.

Let X be the random variable which denotes your winnings in a 1 dollar bet on red in Las Vegas roulette. Then the distribution of X is given by

$$m_X = \begin{pmatrix} -1 & 1 \\ 20/38 & 18/38 \end{pmatrix} ,$$

and one can easily calculate (see Exercise 5) that

$$E(X) \approx -.0526 .$$

We now consider the roulette game in Monte Carlo, and follow the treatment of Sagan.[1] In the roulette game in Monte Carlo there is only one 0. If you bet 1 franc on red and a 0 turns up, then, depending upon the casino, one or more of the following options may be offered:

(a) You get 1/2 of your bet back, and the casino gets the other half of your bet.

(b) Your bet is put "in prison," which we will denote by P_1. If red comes up on the next turn, you get your bet back (but you don't win any money). If black or 0 comes up, you lose your bet.

(c) Your bet is put in prison P_1, as before. If red comes up on the next turn, you get your bet back, and if black comes up on the next turn, then you lose your bet. If a 0 comes up on the next turn, then your bet is put into double prison, which we will denote by P_2. If your bet is in double prison, and if red comes up on the next turn, then your bet is moved back to prison P_1 and the game proceeds as before. If your bet is in double prison, and if black or 0 come up on the next turn, then you lose your bet. We refer the reader to Figure 6.2, where a tree for this option is shown. In this figure, S is the starting position, W means that you win your bet, L means that you lose your bet, and E means that you break even.

[1] H. Sagan, *Markov Chains in Monte Carlo,* Math. Mag., vol. 54, no. 1 (1981), pp. 3-10.

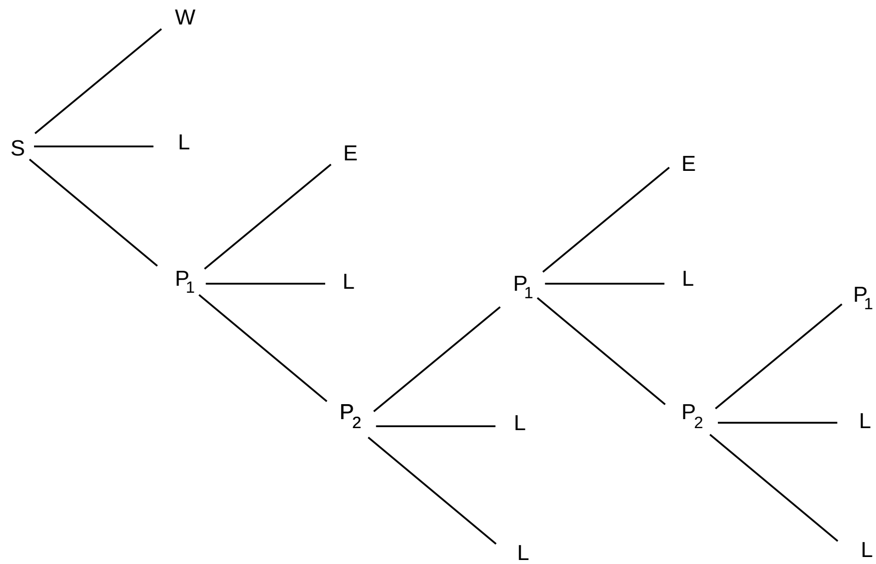

Figure 6.2: Tree for 2-prison Monte Carlo roulette.

It is interesting to compare the expected winnings of a 1 franc bet on red, under each of these three options. We leave the first two calculations as an exercise (see Exercise 37). Suppose that you choose to play alternative (c). The calculation for this case illustrates the way that the early French probabilists worked problems like this.

Suppose you bet on red, you choose alternative (c), and a 0 comes up. Your possible future outcomes are shown in the tree diagram in Figure 6.3. Assume that your money is in the first prison and let x be the probability that you lose your franc. From the tree diagram we see that

$$x = \frac{18}{37} + \frac{1}{37} P(\text{you lose your franc} \mid \text{your franc is in } P_2)\ .$$

Also,

$$P(\text{you lose your franc} \mid \text{your franc is in } P_2) = \frac{19}{37} + \frac{18}{37}x\ .$$

So, we have

$$x = \frac{18}{37} + \frac{1}{37}\Bigl(\frac{19}{37} + \frac{18}{37}x\Bigr)\ .$$

Solving for x, we obtain $x = 685/1351$. Thus, starting at S, the probability that you lose your bet equals

$$\frac{18}{37} + \frac{1}{37}x = \frac{25003}{49987}\ .$$

To find the probability that you win when you bet on red, note that you can only win if red comes up on the first turn, and this happens with probability 18/37. Thus your expected winnings are

$$1 \cdot \frac{18}{37} - 1 \cdot \frac{25003}{49987} = -\frac{687}{49987} \approx -.0137\ .$$

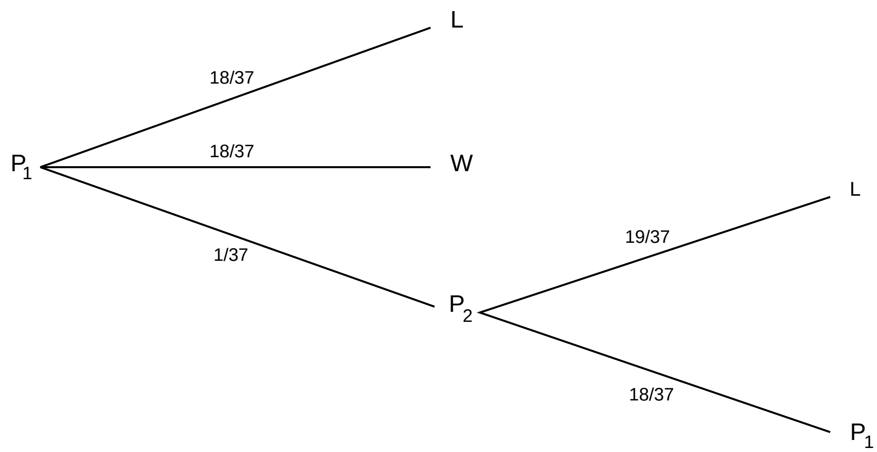

Figure 6.3: Your money is put in prison.

It is interesting to note that the more romantic option (c) is less favorable than option (a) (see Exercise 37).

If you bet 1 dollar on the number 17, then the distribution function for your winnings X is

$$P_X = \begin{pmatrix} -1 & 35 \\ 36/37 & 1/37 \end{pmatrix} ,$$

and the expected winnings are

$$-1 \cdot \frac{36}{37} + 35 \cdot \frac{1}{37} = -\frac{1}{37} \approx -.027 .$$

Thus, at Monte Carlo different bets have different expected values. In Las Vegas almost all bets have the same expected value of $-2/38 = -.0526$ (see Exercises 4 and 5). $\square$

Conditional Expectation

Definition 6.2 If F is any event and X is a random variable with sample space $\Omega = \{x_1, x_2, \ldots\}$, then the *conditional expectation given* F is defined by

$$E(X|F) = \sum_j x_j P(X = x_j|F) .$$

Conditional expectation is used most often in the form provided by the following theorem. $\square$

Theorem 6.5 Let X be a random variable with sample space Ω. If F_1, F_2, $\ldots$, F_r are events such that $F_i \cap F_j = \emptyset$ for $i \neq j$ and $\Omega = \cup_j F_j$, then

$$E(X) = \sum_j E(X|F_j)P(F_j) .$$

Proof. We have

$$\begin{aligned}
\sum_j E(X|F_j)P(F_j) &= \sum_j \sum_k x_k P(X = x_k|F_j)P(F_j) \\
&= \sum_j \sum_k x_k P(X = x_k \text{ and } F_j \text{ occurs}) \\
&= \sum_k \sum_j x_k P(X = x_k \text{ and } F_j \text{ occurs}) \\
&= \sum_k x_k P(X = x_k) \\
&= E(X) \ .
\end{aligned}$$

$\square$

Example 6.14 (Example 6.12 continued) Let T be the number of rolls in a single play of craps. We can think of a single play as a two-stage process. The first stage consists of a single roll of a pair of dice. The play is over if this roll is a 2, 3, 7, 11, or 12. Otherwise, the player's point is established, and the second stage begins. This second stage consists of a sequence of rolls which ends when either the player's point or a 7 is rolled. We record the outcomes of this two-stage experiment using the random variables X and S, where X denotes the first roll, and S denotes the number of rolls in the second stage of the experiment (of course, S is sometimes equal to 0). Note that $T = S + 1$. Then by Theorem 6.5

$$E(T) = \sum_{j=2}^{12} E(T|X = j)P(X = j) \ .$$

If $j = 7$, 11 or 2, 3, 12, then $E(T|X = j) = 1$. If $j = 4, 5, 6, 8, 9,$ or 10, we can use Example 6.4 to calculate the expected value of S. In each of these cases, we continue rolling until we get either a j or a 7. Thus, S is geometrically distributed with parameter p, which depends upon j. If $j = 4$, for example, the value of p is $3/36 + 6/36 = 1/4$. Thus, in this case, the expected number of additional rolls is $1/p = 4$, so $E(T|X = 4) = 1 + 4 = 5$. Carrying out the corresponding calculations for the other possible values of j and using Theorem 6.5 gives

$$\begin{aligned}
E(T) &= 1\Bigl(\frac{12}{36}\Bigr) + \Bigl(1 + \frac{36}{3+6}\Bigr)\Bigl(\frac{3}{36}\Bigr) + \Bigl(1 + \frac{36}{4+6}\Bigr)\Bigl(\frac{4}{36}\Bigr) \\
&\quad + \Bigl(1 + \frac{36}{5+6}\Bigr)\Bigl(\frac{5}{36}\Bigr) + \Bigl(1 + \frac{36}{5+6}\Bigr)\Bigl(\frac{5}{36}\Bigr) \\
&\quad + \Bigl(1 + \frac{36}{4+6}\Bigr)\Bigl(\frac{4}{36}\Bigr) + \Bigl(1 + \frac{36}{3+6}\Bigr)\Bigl(\frac{3}{36}\Bigr) \\
&= \frac{557}{165} \\
&\approx 3.375\ldots \ .
\end{aligned}$$

$\square$

Martingales

We can extend the notion of fairness to a player playing a sequence of games by using the concept of conditional expectation.

Example 6.15 Let S_1, S_2, ..., S_n be Peter's accumulated fortune in playing heads or tails (see Example 1.4). Then

$$E(S_n|S_{n-1} = a, \dots, S_1 = r) = \frac{1}{2}(a+1) + \frac{1}{2}(a-1) = a\ .$$

We note that Peter's expected fortune after the next play is equal to his present fortune. When this occurs, we say the game is *fair*. A fair game is also called a *martingale*. If the coin is biased and comes up heads with probability p and tails with probability $q = 1 - p$, then

$$E(S_n|S_{n-1} = a, \dots, S_1 = r) = p(a+1) + q(a-1) = a + p - q\ .$$

Thus, if $p < q$, this game is unfavorable, and if $p > q$, it is favorable. □

If you are in a casino, you will see players adopting elaborate *systems* of play to try to make unfavorable games favorable. Two such systems, the martingale doubling system and the more conservative Labouchere system, were described in Exercises 1.1.9 and 1.1.10. Unfortunately, such systems cannot change even a fair game into a favorable game.

Even so, it is a favorite pastime of many people to develop systems of play for gambling games and for other games such as the stock market. We close this section with a simple illustration of such a system.

Stock Prices

Example 6.16 Let us assume that a stock increases or decreases in value each day by 1 dollar, each with probability 1/2. Then we can identify this simplified model with our familiar game of heads or tails. We assume that a buyer, Mr. Ace, adopts the following strategy. He buys the stock on the first day at its price V. He then waits until the price of the stock increases by one to $V + 1$ and sells. He then continues to watch the stock until its price falls back to V. He buys again and waits until it goes up to $V + 1$ and sells. Thus he holds the stock in intervals during which it increases by 1 dollar. In each such interval, he makes a profit of 1 dollar. However, we assume that he can do this only for a finite number of trading days. Thus he can lose if, in the last interval that he holds the stock, it does not get back up to $V + 1$; and this is the only way he can lose. In Figure 6.4 we illustrate a typical history if Mr. Ace must stop in twenty days. Mr. Ace holds the stock under his system during the days indicated by broken lines. We note that for the history shown in Figure 6.4, his system nets him a gain of 4 dollars.

We have written a program **StockSystem** to simulate the fortune of Mr. Ace if he uses his sytem over an n-day period. If one runs this program a large number

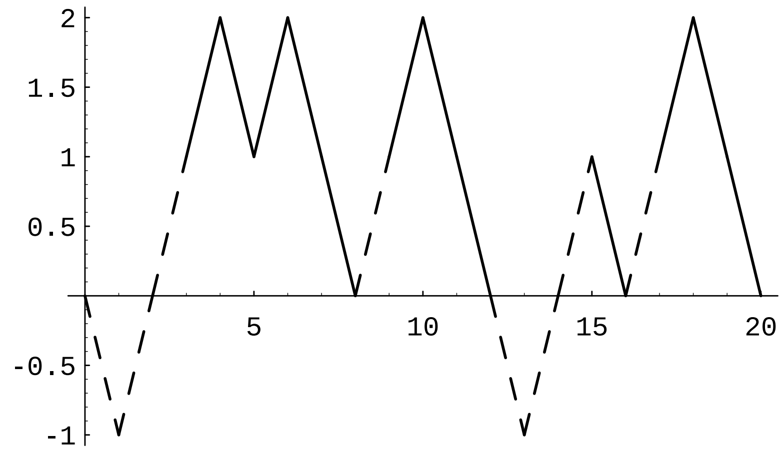

Figure 6.4: Mr. Ace's system.

of times, for $n = 20$, say, one finds that his expected winnings are very close to 0, but the probability that he is ahead after 20 days is significantly greater than 1/2. For small values of n, the exact distribution of winnings can be calculated. The distribution for the case $n = 20$ is shown in Figure 6.5. Using this distribution, it is easy to calculate that the expected value of his winnings is exactly 0. This is another instance of the fact that a fair game (a martingale) remains fair under quite general systems of play.

Although the expected value of his winnings is 0, the probability that Mr. Ace is ahead after 20 days is about .610. Thus, he would be able to tell his friends that his system gives him a better chance of being ahead than that of someone who simply buys the stock and holds it, if our simple random model is correct. There have been a number of studies to determine how random the stock market is. □

Historical Remarks

With the Law of Large Numbers to bolster the frequency interpretation of probability, we find it natural to justify the definition of expected value in terms of the average outcome over a large number of repetitions of the experiment. The concept of expected value was used before it was formally defined; and when it was used, it was considered not as an average value but rather as the appropriate value for a gamble. For example recall, from the Historical Remarks section of Chapter 1, Section 1.2, Pascal's way of finding the value of a three-game series that had to be called off before it is finished.

Pascal first observed that if each player has only one game to win, then the stake of 64 pistoles should be divided evenly. Then he considered the case where one player has won two games and the other one.

> Then consider, Sir, if the first man wins, he gets 64 pistoles, if he loses he gets 32. Thus if they do not wish to risk this last game, but wish

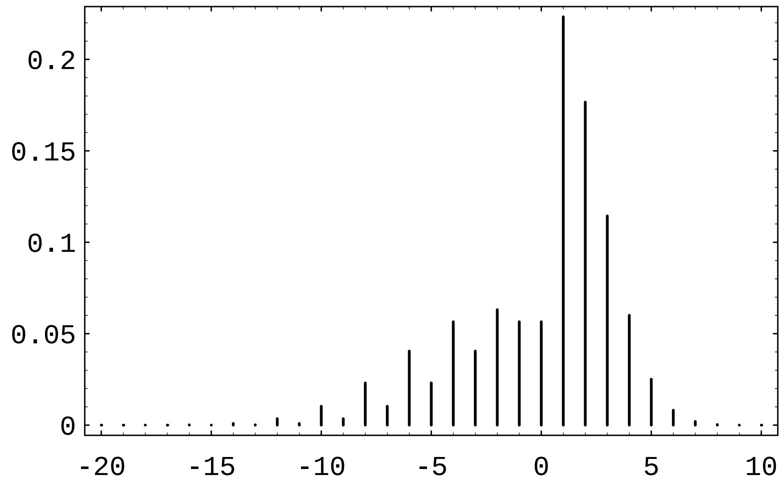

Figure 6.5: Winnings distribution for $n = 20$.

> to separate without playing it, the first man must say: "I am certain to get 32 pistoles, even if I lose I still get them; but as for the other 32 pistoles, perhaps I will get them, perhaps you will get them, the chances are equal. Let us then divide these 32 pistoles in half and give one half to me as well as my 32 which are mine for sure." He will then have 48 pistoles and the other 16.[2]

Note that Pascal reduced the problem to a symmetric bet in which each player gets the same amount and takes it as obvious that in this case the stakes should be divided equally.

The first systematic study of expected value appears in Huygens' book. Like Pascal, Huygens find the value of a gamble by assuming that the answer is obvious for certain symmetric situations and uses this to deduce the expected for the general situation. He does this in steps. His first proposition is

> Prop. I. If I expect a or b, either of which, with equal probability, may fall to me, then my Expectation is worth $(a+b)/2$, that is, the half Sum of a and b.[3]

Huygens proved this as follows: Assume that two player A and B play a game in which each player puts up a stake of $(a+b)/2$ with an equal chance of winning the total stake. Then the value of the game to each player is $(a+b)/2$. For example, if the game had to be called off clearly each player should just get back his original stake. Now, by symmetry, this value is not changed if we add the condition that the winner of the game has to pay the loser an amount b as a consolation prize. Then for player A the value is still $(a+b)/2$. But what are his possible outcomes

[2]Quoted in F. N. David, *Games, Gods and Gambling* (London: Griffin, 1962), p. 231.

[3]C. Huygens, *Calculating in Games of Chance,* translation attributed to John Arbuthnot (London, 1692), p. 34.

for the modified game? If he wins he gets the total stake $a + b$ and must pay B an amount b so ends up with a. If he loses he gets an amount b from player B. Thus player A wins a or b with equal chances and the value to him is $(a + b)/2$.

Huygens illustrated this proof in terms of an example. If you are offered a game in which you have an equal chance of winning 2 or 8, the expected value is 5, since this game is equivalent to the game in which each player stakes 5 and agrees to pay the loser 3 — a game in which the value is obviously 5.

Huygens' second proposition is

> Prop. II. If I expect a, b, or c, either of which, with equal facility, may happen, then the Value of my Expectation is $(a + b + c)/3$, or the third of the Sum of a, b, and c.[4]

His argument here is similar. Three players, A, B, and C, each stake

$$(a + b + c)/3$$

in a game they have an equal chance of winning. The value of this game to player A is clearly the amount he has staked. Further, this value is not changed if A enters into an agreement with B that if one of them wins he pays the other a consolation prize of b and with C that if one of them wins he pays the other a consolation prize of c. By symmetry these agreements do not change the value of the game. In this modified game, if A wins he wins the total stake $a + b + c$ minus the consolation prizes $b + c$ giving him a final winning of a. If B wins, A wins b and if C wins, A wins c. Thus A finds himself in a game with value $(a + b + c)/3$ and with outcomes a, b, and c occurring with equal chance. This proves Proposition II.

More generally, this reasoning shows that if there are n outcomes

$$a_1,\ a_2,\ \ldots,\ a_n\ ,$$

all occurring with the same probability, the expected value is

$$\frac{a_1 + a_2 + \cdots + a_n}{n}\ .$$

In his third proposition Huygens considered the case where you win a or b but with unequal probabilities. He assumed there are p chances of winning a, and q chances of winning b, all having the same probability. He then showed that the expected value is

$$E = \frac{p}{p+q} \cdot a + \frac{q}{p+q} \cdot b\ .$$

This follows by considering an equivalent gamble with $p + q$ outcomes all occurring with the same probability and with a payoff of a in p of the outcomes and b in q of the outcomes. This allowed Huygens to compute the expected value for experiments with unequal probabilities, at least when these probablities are rational numbers.

Thus, instead of defining the expected value as a weighted average, Huygens assumed that the expected value of certain symmetric gambles are known and deduced the other values from these. Although this requires a good deal of clever

[4]ibid., p. 35.

manipulation, Huygens ended up with values that agree with those given by our modern definition of expected value. One advantage of this method is that it gives a justification for the expected value in cases where it is not reasonable to assume that you can repeat the experiment a large number of times, as for example, in betting that at least two presidents died on the same day of the year. (In fact, three did; all were signers of the Declaration of Independence, and all three died on July 4.)

In his book, Huygens calculated the expected value of games using techniques similar to those which we used in computing the expected value for roulette at Monte Carlo. For example, his proposition XIV is:

> Prop. XIV. If I were playing with another by turns, with two Dice, on this Condition, that if I throw 7 I gain, and if he throws 6 he gains allowing him the first Throw: To find the proportion of my Hazard to his.[5]

A modern description of this game is as follows. Huygens and his opponent take turns rolling a die. The game is over if Huygens rolls a 7 or his opponent rolls a 6. His opponent rolls first. What is the probability that Huygens wins the game?

To solve this problem Huygens let x be his chance of winning when his opponent threw first and y his chance of winning when he threw first. Then on the first roll his opponent wins on 5 out of the 36 possibilities. Thus,

$$x = \frac{31}{36} \cdot y \ .$$

But when Huygens rolls he wins on 6 out of the 36 possible outcomes, and in the other 30, he is led back to where his chances are x. Thus

$$y = \frac{6}{36} + \frac{30}{36} \cdot x \ .$$

From these two equations Huygens found that $x = 31/61$.

Another early use of expected value appeared in Pascal's argument to show that a rational person should believe in the existence of God.[6] Pascal said that we have to make a wager whether to believe or not to believe. Let p denote the probability that God does not exist. His discussion suggests that we are playing a game with two strategies, believe and not believe, with payoffs as shown in Table 6.4.

Here $-u$ represents the cost to you of passing up some worldly pleasures as a consequence of believing that God exists. If you do not believe, and God is a vengeful God, you will lose x. If God exists and you do believe you will gain v. Now to determine which strategy is best you should compare the two expected values

$$p(-u) + (1-p)v \qquad \text{and} \qquad p0 + (1-p)(-x),$$

[5] ibid., p. 47.

[6] Quoted in I. Hacking, *The Emergence of Probability* (Cambridge: Cambridge Univ. Press, 1975).

	God does not exist	God exists
	p	$1-p$
believe	$-u$	v
not believe	0	$-x$

Table 6.4: Payoffs.

Age	Survivors
0	100
6	64
16	40
26	25
36	16
46	10
56	6
66	3
76	1

Table 6.5: Graunt's mortality data.

and choose the larger of the two. In general, the choice will depend upon the value of p. But Pascal assumed that the value of v is infinite and so the strategy of believing is best no matter what probability you assign for the existence of God. This example is considered by some to be the beginning of decision theory. Decision analyses of this kind appear today in many fields, and, in particular, are an important part of medical diagnostics and corporate business decisions.

Another early use of expected value was to decide the price of annuities. The study of statistics has its origins in the use of the bills of mortality kept in the parishes in London from 1603. These records kept a weekly tally of christenings and burials. From these John Graunt made estimates for the population of London and also provided the first mortality data,[7] shown in Table 6.5.

As Hacking observes, Graunt apparently constructed this table by assuming that after the age of 6 there is a constant probability of about 5/8 of surviving for another decade.[8] For example, of the 64 people who survive to age 6, 5/8 of 64 or 40 survive to 16, 5/8 of these 40 or 25 survive to 26, and so forth. Of course, he rounded off his figures to the nearest whole person.

Clearly, a constant mortality rate cannot be correct throughout the whole range, and later tables provided by Halley were more realistic in this respect.[9]

[7]ibid., p. 108.
[8]ibid., p. 109.
[9]E. Halley, "An Estimate of The Degrees of Mortality of Mankind," *Phil. Trans. Royal. Soc.,*

A *terminal annuity* provides a fixed amount of money during a period of n years. To determine the price of a terminal annuity one needs only to know the appropriate interest rate. A *life annuity* provides a fixed amount during each year of the buyer's life. The appropriate price for a life annuity is the expected value of the terminal annuity evaluated for the random lifetime of the buyer. Thus, the work of Huygens in introducing expected value and the work of Graunt and Halley in determining mortality tables led to a more rational method for pricing annuities. This was one of the first serious uses of probability theory outside the gambling houses.

Although expected value plays a role now in every branch of science, it retains its importance in the casino. In 1962, Edward Thorp's book *Beat the Dealer*[10] provided the reader with a strategy for playing the popular casino game of blackjack that would assure the player a positive expected winning. This book forevermore changed the belief of the casinos that they could not be beat.

Exercises

1 A card is drawn at random from a deck consisting of cards numbered 2 through 10. A player wins 1 dollar if the number on the card is odd and loses 1 dollar if the number if even. What is the expected value of his winnings?

2 A card is drawn at random from a deck of playing cards. If it is red, the player wins 1 dollar; if it is black, the player loses 2 dollars. Find the expected value of the game.

3 In a class there are 20 students: 3 are 5' 6", 5 are 5'8", 4 are 5'10", 4 are 6', and 4 are 6' 2". A student is chosen at random. What is the student's expected height?

4 In Las Vegas the roulette wheel has a 0 and a 00 and then the numbers 1 to 36 marked on equal slots; the wheel is spun and a ball stops randomly in one slot. When a player bets 1 dollar on a number, he receives 36 dollars if the ball stops on this number, for a net gain of 35 dollars; otherwise, he loses his dollar bet. Find the expected value for his winnings.

5 In a second version of roulette in Las Vegas, a player bets on red or black. Half of the numbers from 1 to 36 are red, and half are black. If a player bets a dollar on black, and if the ball stops on a black number, he gets his dollar back and another dollar. If the ball stops on a red number or on 0 or 00 he loses his dollar. Find the expected winnings for this bet.

6 A die is rolled twice. Let X denote the sum of the two numbers that turn up, and Y the difference of the numbers (specifically, the number on the first roll minus the number on the second). Show that $E(XY) = E(X)E(Y)$. Are X and Y independent?

---

vol. 17 (1693), pp. 596–610; 654–656.

[10] E. Thorp, *Beat the Dealer* (New York: Random House, 1962).

***7** Show that, if X and Y are random variables taking on only two values each, and if $E(XY) = E(X)E(Y)$, then X and Y are independent.

8 A royal family has children until it has a boy or until it has three children, whichever comes first. Assume that each child is a boy with probability 1/2. Find the expected number of boys in this royal family and the expected number of girls.

9 If the first roll in a game of craps is neither a natural nor craps, the player can make an additional bet, equal to his original one, that he will make his point before a seven turns up. If his point is four or ten he is paid off at 2 : 1 odds; if it is a five or nine he is paid off at odds 3 : 2; and if it is a six or eight he is paid off at odds 6 : 5. Find the player's expected winnings if he makes this additional bet when he has the opportunity.

10 In Example 6.16 assume that Mr. Ace decides to buy the stock and hold it until it goes up 1 dollar and then sell and not buy again. Modify the program **StockSystem** to find the distribution of his profit under this system after a twenty-day period. Find the expected profit and the probability that he comes out ahead.

11 On September 26, 1980, the *New York Times* reported that a mysterious stranger strode into a Las Vegas casino, placed a single bet of 777,000 dollars on the "don't pass" line at the crap table, and walked away with more than 1.5 million dollars. In the "don't pass" bet, the bettor is essentially betting with the house. An exception occurs if the roller rolls a 12 on the first roll. In this case, the roller loses and the "don't pass" better just gets back the money bet instead of winning. Show that the "don't pass" bettor has a more favorable bet than the roller.

12 Recall that in the *martingale doubling system* (see Exercise 1.1.10), the player doubles his bet each time he loses. Suppose that you are playing roulette in a *fair casino* where there are no 0's, and you bet on red each time. You then win with probability 1/2 each time. Assume that you enter the casino with 100 dollars, start with a 1-dollar bet and employ the martingale system. You stop as soon as you have won one bet, or in the unlikely event that black turns up six times in a row so that you are down 63 dollars and cannot make the required 64-dollar bet. Find your expected winnings under this system of play.

13 You have 80 dollars and play the following game. An urn contains two white balls and two black balls. You draw the balls out one at a time without replacement until all the balls are gone. On each draw, you bet half of your present fortune that you will draw a white ball. What is your expected final fortune?

14 In the hat check problem (see Example 3.12), it was assumed that N people check their hats and the hats are handed back at random. Let $X_j = 1$ if the

jth person gets his or her hat and 0 otherwise. Find $E(X_j)$ and $E(X_j \cdot X_k)$ for j not equal to k. Are X_j and X_k independent?

15 A box contains two gold balls and three silver balls. You are allowed to choose successively balls from the box at random. You win 1 dollar each time you draw a gold ball and lose 1 dollar each time you draw a silver ball. After a draw, the ball is not replaced. Show that, if you draw until you are ahead by 1 dollar or until there are no more gold balls, this is a favorable game.

16 Gerolamo Cardano in his book, *The Gambling Scholar,* written in the early 1500s, considers the following carnival game. There are six dice. Each of the dice has five blank sides. The sixth side has a number between 1 and 6—a different number on each die. The six dice are rolled and the player wins a prize depending on the total of the numbers which turn up.

(a) Find, as Cardano did, the expected total without finding its distribution.

(b) Large prizes were given for large totals with a modest fee to play the game. Explain why this could be done.

17 Let X be the first time that a *failure* occurs in an infinite sequence of Bernoulli trials with probability p for success. Let $p_k = P(X = k)$ for $k = 1, 2, \ldots$. Show that $p_k = p^{k-1}q$ where $q = 1 - p$. Show that $\sum_k p_k = 1$. Show that $E(X) = 1/q$. What is the expected number of tosses of a coin required to obtain the first tail?

18 Exactly one of six similar keys opens a certain door. If you try the keys, one after another, what is the expected number of keys that you will have to try before success?

19 A multiple choice exam is given. A problem has four possible answers, and exactly one answer is correct. The student is allowed to choose a subset of the four possible answers as his answer. If his chosen subset contains the correct answer, the student receives three points, but he loses one point for each wrong answer in his chosen subset. Show that if he just guesses a subset uniformly and randomly his expected score is zero.

20 You are offered the following game to play: a fair coin is tossed until heads turns up for the first time (see Example 6.3). If this occurs on the first toss you receive 2 dollars, if it occurs on the second toss you receive $2^2 = 4$ dollars and, in general, if heads turns up for the first time on the nth toss you receive 2^n dollars.

(a) Show that the expected value of your winnings does not exist (i.e., is given by a divergent sum) for this game. Does this mean that this game is favorable no matter how much you pay to play it?

(b) Assume that you only receive 2^{10} dollars if any number greater than or equal to ten tosses are required to obtain the first head. Show that your expected value for this modified game is finite and find its value.

(c) Assume that you pay 10 dollars for each play of the original game. Write a program to simulate 100 plays of the game and see how you do.

(d) Now assume that the utility of n dollars is $\sqrt{n}$. Write an expression for the expected utility of the payment, and show that this expression has a finite value. Estimate this value. Repeat this exercise for the case that the utility function is $\log(n)$.

21 Let X be a random variable which is Poisson distributed with parameter λ. Show that $E(X) = \lambda$. *Hint*: Recall that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots .$$

22 Recall that in Exercise 1.1.14, we considered a town with two hospitals. In the large hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. We were interested in guessing which hospital would have on the average the largest number of days with the property that more than 60 percent of the children born on that day are boys. For each hospital find the expected number of days in a year that have the property that more than 60 percent of the children born on that day were boys.

23 An insurance company has 1,000 policies on men of age 50. The company estimates that the probability that a man of age 50 dies within a year is .01. Estimate the number of claims that the company can expect from beneficiaries of these men within a year.

24 Using the life table for 1981 in Appendix C, write a program to compute the expected lifetime for males and females of each possible age from 1 to 85. Compare the results for males and females. Comment on whether life insurance should be priced differently for males and females.

***25** A deck of ESP cards consists of 20 cards each of two types: say ten stars, ten circles (normally there are five types). The deck is shuffled and the cards turned up one at a time. You, the alleged percipient, are to name the symbol on each card *before* it is turned up.

Suppose that you are really just guessing at the cards. If you do not get to see each card after you have made your guess, then it is easy to calculate the expected number of correct guesses, namely ten.

If, on the other hand, you are guessing with information, that is, if you see each card after your guess, then, of course, you might expect to get a higher score. This is indeed the case, but calculating the correct expectation is no longer easy.

But it is easy to do a computer simulation of this guessing with information, so we can get a good idea of the expectation by simulation. (This is similar to the way that skilled blackjack players make blackjack into a favorable game by observing the cards that have already been played. See Exercise 29.)

(a) First, do a simulation of guessing without information, repeating the experiment at least 1000 times. Estimate the expected number of correct answers and compare your result with the theoretical expectation.

(b) What is the best strategy for guessing with information?

(c) Do a simulation of guessing with information, using the strategy in (b). Repeat the experiment at least 1000 times, and estimate the expectation in this case.

(d) Let S be the number of stars and C the number of circles in the deck. Let $h(S,C)$ be the expected winnings using the optimal guessing strategy in (b). Show that $h(S,C)$ satisfies the recursion relation

$$h(S,C) = \frac{S}{S+C}h(S-1,C) + \frac{C}{S+C}h(S,C-1) + \frac{\max(S,C)}{S+C}\ ,$$

and $h(0,0) = h(-1,0) = h(0,-1) = 0$. Using this relation, write a program to compute $h(S,C)$ and find $h(10,10)$. Compare the computed value of $h(10,10)$ with the result of your simulation in (c). For more about this exercise and Exercise 26 see Diaconis and Graham.[11]

***26** Consider the ESP problem as described in Exercise 25. You are again guessing with information, and you are using the optimal guessing strategy of guessing *star* if the remaining deck has more stars, *circle* if more circles, and tossing a coin if the number of stars and circles are equal. Assume that $S \geq C$, where S is the number of stars and C the number of circles.

We can plot the results of a typical game on a graph, where the horizontal axis represents the number of steps and the vertical axis represents the *difference* between the number of stars and the number of circles that have been turned up. A typical game is shown in Figure 6.6. In this particular game, the order in which the cards were turned up is (C,S,S,S,S,C,C,S,S,C). Thus, in this particular game, there were six stars and four circles in the deck. This means, in particular, that every game played with this deck would have a graph which ends at the point $(10,2)$. We define the line L to be the horizontal line which goes through the ending point on the graph (so its vertical coordinate is just the difference between the number of stars and circles in the deck).

(a) Show that, when the random walk is below the line L, the player guesses right when the graph goes up (star is turned up) and, when the walk is above the line, the player guesses right when the walk goes down (circle turned up). Show from this property that the subject is sure to have at least S correct guesses.

(b) When the walk is at a point (x,x) *on* the line L the number of stars and circles remaining is the same, and so the subject tosses a coin. Show that

[11] P. Diaconis and R. Graham, "The Analysis of Sequential Experiments with Feedback to Subjects," *Annals of Statistics,* vol. 9 (1981), pp. 3–23.

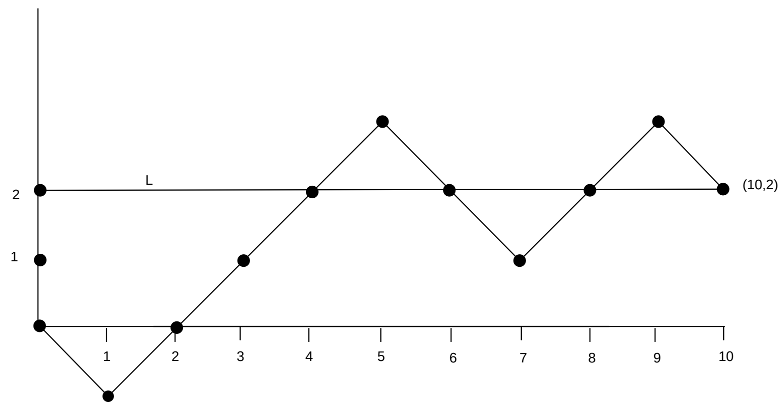

Figure 6.6: Random walk for ESP.

the probability that the walk reaches (x, x) is

$$\frac{\binom{S}{x}\binom{C}{x}}{\binom{S+C}{2x}}\ .$$

Hint: The outcomes of $2x$ cards is a hypergeometric distribution (see Section 5.1).

(c) Using the results of (a) and (b) show that the expected number of correct guesses under intelligent guessing is

$$S + \sum_{x=1}^{C} \frac{1}{2} \frac{\binom{S}{x}\binom{C}{x}}{\binom{S+C}{2x}}\ .$$

27 It has been said[12] that a Dr. B. Muriel Bristol declined a cup of tea stating that she preferred a cup into which milk had been poured first. The famous statistician R. A. Fisher carried out a test to see if she could tell whether milk was put in before or after the tea. Assume that for the test Dr. Bristol was given eight cups of tea—four in which the milk was put in before the tea and four in which the milk was put in after the tea.

(a) What is the expected number of correct guesses the lady would make if she had no information after each test and was just guessing?

(b) Using the result of Exercise 26 find the expected number of correct guesses if she was told the result of each guess and used an optimal guessing strategy.

28 In a popular computer game the computer picks an integer from 1 to n at random. The player is given k chances to guess the number. After each guess the computer responds "correct," "too small," or "too big."

[12] J. F. Box, *R. A. Fisher, The Life of a Scientist* (New York: John Wiley and Sons, 1978).

(a) Show that if $n \leq 2^k - 1$, then there is a strategy that guarantees you will correctly guess the number in k tries.

(b) Show that if $n \geq 2^k - 1$, there is a strategy that assures you of identifying one of $2^k - 1$ numbers and hence gives a probability of $(2^k - 1)/n$ of winning. Why is this an optimal strategy? Illustrate your result in terms of the case $n = 9$ and $k = 3$.

29 In the casino game of blackjack the dealer is dealt two cards, one face up and one face down, and each player is dealt two cards, both face down. If the dealer is showing an ace the player can look at his down cards and then make a bet called an *insurance* bet. (Expert players will recognize why it is called insurance.) If you make this bet you will win the bet if the dealer's second card is a *ten card*: namely, a ten, jack, queen, or king. If you win, you are paid twice your insurance bet; otherwise you lose this bet. Show that, if the only cards you can see are the dealer's ace and your two cards and if your cards are not ten cards, then the insurance bet is an unfavorable bet. Show, however, that if you are playing two hands simultaneously, and you have no ten cards, then it is a favorable bet. (Thorp[13] has shown that the game of blackjack is favorable to the player if he or she can keep good enough track of the cards that have been played.)

30 Assume that, every time you buy a box of Wheaties, you receive a picture of one of the n players for the New York Yankees (see Exercise 3.2.34). Let X_k be the number of additional boxes you have to buy, after you have obtained $k-1$ different pictures, in order to obtain the next new picture. Thus $X_1 = 1$, X_2 is the number of boxes bought after this to obtain a picture different from the first pictured obtained, and so forth.

(a) Show that X_k has a geometric distribution with $p = (n - k + 1)/n$.

(b) Simulate the experiment for a team with 26 players (25 would be more accurate but we want an even number). Carry out a number of simulations and estimate the expected time required to get the first 13 players and the expected time to get the second 13. How do these expectations compare?

(c) Show that, if there are $2n$ players, the expected time to get the first half of the players is

$$2n\left(\frac{1}{2n} + \frac{1}{2n-1} + \cdots + \frac{1}{n+1}\right) ,$$

and the expected time to get the second half is

$$2n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + 1\right) .$$

[13] E. Thorp, *Beat the Dealer* (New York: Random House, 1962).

(d) In Example 6.11 we stated that

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \sim \log n + .5772 + \frac{1}{2n} \ .$$

Use this to estimate the expression in (c). Compare these estimates with the exact values and also with your estimates obtained by simulation for the case $n = 26$.

***31** (Feller[14]) A large number, N, of people are subjected to a blood test. This can be administered in two ways: (1) Each person can be tested separately, in this case N test are required, (2) the blood samples of k persons can be pooled and analyzed together. If this test is *negative,* this one test suffices for the k people. If the test is *positive,* each of the k persons must be tested separately, and in all, $k + 1$ tests are required for the k people. Assume that the probability p that a test is positive is the same for all people and that these events are independent.

(a) Find the probability that the test for a pooled sample of k people will be positive.

(b) What is the expected value of the number X of tests necessary under plan (2)? (Assume that N is divisible by k.)

(c) For small p, show that the value of k which will minimize the expected number of tests under the second plan is approximately $1/\sqrt{p}$.

32 Write a program to add random numbers chosen from $[0, 1]$ until the first time the sum is greater than one. Have your program repeat this experiment a number of times to estimate the expected number of selections necessary in order that the sum of the chosen numbers first exceeds 1. On the basis of your experiments, what is your estimate for this number?

***33** The following related discrete problem also gives a good clue for the answer to Exercise 32. Randomly select with replacement t_1, t_2, ..., t_r from the set $(1/n, 2/n, \ldots, n/n)$. Let X be the smallest value of r satisfying

$$t_1 + t_2 + \cdots + t_r > 1 \ .$$

Then $E(X) = (1 + 1/n)^n$. To prove this, we can just as well choose t_1, t_2, ..., t_r randomly with replacement from the set $(1, 2, \ldots, n)$ and let X be the smallest value of r for which

$$t_1 + t_2 + \cdots + t_r > n \ .$$

(a) Use Exercise 3.2.36 to show that

$$P(X \geq j + 1) = \binom{n}{j} \left(\frac{1}{n}\right)^j \ .$$

[14] W. Feller, *Introduction to Probability Theory and Its Applications,* 3rd ed., vol. 1 (New York: John Wiley and Sons, 1968), p. 240.

(b) Show that

$$E(X) = \sum_{j=0}^{n} P(X \geq j+1) \ .$$

(c) From these two facts, find an expression for $E(X)$. This proof is due to Harris Schultz.[15]

***34** (Banach's Matchbox[16]) A man carries in each of his two front pockets a box of matches originally containing N matches. Whenever he needs a match, he chooses a pocket at random and removes one from that box. One day he reaches into a pocket and finds the box empty.

(a) Let p_r denote the probability that the other pocket contains r matches. Define a sequence of *counter* random variables as follows: Let $X_i = 1$ if the ith draw is from the left pocket, and 0 if it is from the right pocket. Interpret p_r in terms of $S_n = X_1 + X_2 + \cdots + X_n$. Find a binomial expression for p_r.

(b) Write a computer program to compute the p_r, as well as the probability that the other pocket contains at least r matches, for $N = 100$ and r from 0 to 50.

(c) Show that $(N-r)p_r = (1/2)(2N+1)p_{r+1} - (1/2)(r+1)p_{r+1}$.

(d) Evaluate $\sum_r p_r$.

(e) Use (c) and (d) to determine the expectation E of the distribution $\{p_r\}$.

(f) Use Stirling's formula to obtain an approximation for E. How many matches must each box contain to ensure a value of about 13 for the expectation E? (Take $\pi = 22/7$.)

35 A coin is tossed until the first time a head turns up. If this occurs on the nth toss and n is odd you win $2^n/n$, but if n is even then you lose $2^n/n$. Then if your expected winnings exist they are given by the convergent series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

called the alternating *harmonic series.* It is tempting to say that this should be the expected value of the experiment. Show that if we were to do this, the expected value of an experiment would depend upon the order in which the outcomes are listed.

36 Suppose we have an urn containing c yellow balls and d green balls. We draw k balls, without replacement, from the urn. Find the expected number of yellow balls drawn. *Hint*: Write the number of yellow balls drawn as the sum of c random variables.

---

[15] H. Schultz, "An Expected Value Problem," *Two-Year Mathematics Journal,* vol. 10, no. 4 (1979), pp. 277–78.

[16] W. Feller, *Introduction to Probability Theory,* vol. 1, p. 166.

37 The reader is referred to Example 6.13 for an explanation of the various options available in Monte Carlo roulette.

(a) Compute the expected winnings of a 1 franc bet on red under option (a).

(b) Repeat part (a) for option (b).

(c) Compare the expected winnings for all three options.

***38** (from Pittel[17]) Telephone books, n in number, are kept in a stack. The probability that the book numbered i (where $1 \leq i \leq n$) is consulted for a given phone call is $p_i > 0$, where the p_i's sum to 1. After a book is used, it is placed at the top of the stack. Assume that the calls are independent and evenly spaced, and that the system has been employed indefinitely far into the past. Let d_i be the average depth of book i in the stack. Show that $d_i \leq d_j$ whenever $p_i \geq p_j$. Thus, on the average, the more popular books have a tendency to be closer to the top of the stack. *Hint*: Let p_{ij} denote the probability that book i is above book j. Show that $p_{ij} = p_{ij}(1 - p_j) + p_{ji}p_i$.

***39** (from Propp[18]) In the previous problem, let P be the probability that at the present time, each book is in its proper place, i.e., book i is ith from the top. Find a formula for P in terms of the p_i's. In addition, find the least upper bound on P, if the p_i's are allowed to vary. *Hint*: First find the probability that book 1 is in the right place. Then find the probability that book 2 is in the right place, given that book 1 is in the right place. Continue.

***40** (from H. Shultz and B. Leonard[19]) A sequence of random numbers in $[0, 1)$ is generated until the sequence is no longer monotone increasing. The numbers are chosen according to the uniform distribution. What is the expected length of the sequence? (In calculating the length, the term that destroys monotonicity is included.) *Hint*: Let $a_1,\ a_2,\ \ldots$ be the sequence and let X denote the length of the sequence. Then

$$P(X > k) = P(a_1 < a_2 < \cdots < a_k)\ ,$$

and the probability on the right-hand side is easy to calculate. Furthermore, one can show that

$$E(X) = 1 + P(X > 1) + P(X > 2) + \cdots\ .$$

41 Let T be the random variable that counts the number of 2-unshuffles performed on an n-card deck until all of the labels on the cards are distinct. This random variable was discussed in Section 3.3. Using Equation 3.4 in that section, together with the formula

$$E(T) = \sum_{s=0}^{\infty} P(T > s)$$

[17] B. Pittel, Problem #1195, *Mathematics Magazine,* vol. 58, no. 3 (May 1985), pg. 183.

[18] J. Propp, Problem #1159, *Mathematics Magazine* vol. 57, no. 1 (Feb. 1984), pg. 50.

[19] H. Shultz and B. Leonard, “Unexpected Occurrences of the Number e,” *Mathematics Magazine* vol. 62, no. 4 (October, 1989), pp. 269-271.

that was proved in Exercise 33, show that

$$E(T) = \sum_{s=0}^{\infty} \left(1 - \binom{2^s}{n} \frac{n!}{2^{sn}}\right) .$$

Show that for $n = 52$, this expression is approximately equal to 11.7. (As was stated in Chapter 3, this means that on the average, almost 12 riffle shuffles of a 52-card deck are required in order for the process to be considered random.)

6.2 Variance of Discrete Random Variables

The usefulness of the expected value as a prediction for the outcome of an experiment is increased when the outcome is not likely to deviate too much from the expected value. In this section we shall introduce a measure of this deviation, called the variance.

Variance

Definition 6.3 Let X be a numerically valued random variable with expected value $\mu = E(X)$. Then the *variance* of X, denoted by $V(X)$, is

$$V(X) = E((X - \mu)^2) .$$

□

Note that, by Theorem 6.1, $V(X)$ is given by

$$V(X) = \sum_x (x - \mu)^2 m(x) , \tag{6.1}$$

where m is the distribution function of X.

Standard Deviation

The *standard deviation* of X, denoted by $D(X)$, is $D(X) = \sqrt{V(X)}$. We often write σ for $D(X)$ and σ^2 for $V(X)$.

Example 6.17 Consider one roll of a die. Let X be the number that turns up. To find $V(X)$, we must first find the expected value of X. This is

$$\begin{aligned} \mu &= E(X) = 1\Big(\frac{1}{6}\Big) + 2\Big(\frac{1}{6}\Big) + 3\Big(\frac{1}{6}\Big) + 4\Big(\frac{1}{6}\Big) + 5\Big(\frac{1}{6}\Big) + 6\Big(\frac{1}{6}\Big) \\ &= \frac{7}{2} . \end{aligned}$$

To find the variance of X, we form the new random variable $(X - \mu)^2$ and compute its expectation. We can easily do this using the following table.

x	$m(x)$	$(x-7/2)^2$
1	1/6	25/4
2	1/6	9/4
3	1/6	1/4
4	1/6	1/4
5	1/6	9/4
6	1/6	25/4

Table 6.6: Variance calculation.

From this table we find $E((X-\mu)^2)$ is

$$\begin{aligned} V(X) &= \frac{1}{6}\left(\frac{25}{4}+\frac{9}{4}+\frac{1}{4}+\frac{1}{4}+\frac{9}{4}+\frac{25}{4}\right) \\ &= \frac{35}{12}\ , \end{aligned}$$

and the standard deviation $D(X) = \sqrt{35/12} \approx 1.707$. □

Calculation of Variance

We next prove a theorem that gives us a useful alternative form for computing the variance.

Theorem 6.6 If X is any random variable with $E(X) = \mu$, then

$$V(X) = E(X^2) - \mu^2\ .$$

Proof. We have

$$\begin{aligned} V(X) &= E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2\ . \end{aligned}$$

□

Using Theorem 6.6, we can compute the variance of the outcome of a roll of a die by first computing

$$\begin{aligned} E(X^2) &= 1\Bigl(\frac{1}{6}\Bigr) + 4\Bigl(\frac{1}{6}\Bigr) + 9\Bigl(\frac{1}{6}\Bigr) + 16\Bigl(\frac{1}{6}\Bigr) + 25\Bigl(\frac{1}{6}\Bigr) + 36\Bigl(\frac{1}{6}\Bigr) \\ &= \frac{91}{6}\ , \end{aligned}$$

and,

$$V(X) = E(X^2) - \mu^2 = \frac{91}{6} - \Bigl(\frac{7}{2}\Bigr)^2 = \frac{35}{12}\ ,$$

in agreement with the value obtained directly from the definition of $V(X)$.

Properties of Variance

The variance has properties very different from those of the expectation. If c is any constant, $E(cX) = cE(X)$ and $E(X + c) = E(X) + c$. These two statements imply that the expectation is a linear function. However, the variance is not linear, as seen in the next theorem.

Theorem 6.7 If X is any random variable and c is any constant, then

$$V(cX) = c^2V(X)$$

and

$$V(X + c) = V(X)\ .$$

Proof. Let $\mu = E(X)$. Then $E(cX) = c\mu$, and

$$\begin{aligned} V(cX) &= E((cX - c\mu)^2) = E(c^2(X - \mu)^2) \\ &= c^2E((X - \mu)^2) = c^2V(X)\ . \end{aligned}$$

To prove the second assertion, we note that, to compute $V(X + c)$, we would replace x by $x+c$ and μ by $\mu+c$ in Equation 6.1. Then the c's would cancel, leaving $V(X)$. $\square$

We turn now to some general properties of the variance. Recall that if X and Y are any two random variables, $E(X+Y) = E(X)+E(Y)$. This is not always true for the case of the variance. For example, let X be a random variable with $V(X) \neq 0$, and define $Y = -X$. Then $V(X) = V(Y)$, so that $V(X) + V(Y) = 2V(X)$. But $X + Y$ is always 0 and hence has variance 0. Thus $V(X + Y) \neq V(X) + V(Y)$.

In the important case of mutually independent random variables, however, *the variance of the sum is the sum of the variances.*

Theorem 6.8 Let X and Y be two *independent* random variables. Then

$$V(X + Y) = V(X) + V(Y)\ .$$

Proof. Let $E(X) = a$ and $E(Y) = b$. Then

$$\begin{aligned} V(X + Y) &= E((X + Y)^2) - (a + b)^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - a^2 - 2ab - b^2\ . \end{aligned}$$

Since X and Y are independent, $E(XY) = E(X)E(Y) = ab$. Thus,

$$V(X + Y) = E(X^2) - a^2 + E(Y^2) - b^2 = V(X) + V(Y)\ .$$

$\square$

It is easy to extend this proof, by mathematical induction, to show that *the variance of the sum of any number of mutually independent random variables is the sum of the individual variances.* Thus we have the following theorem.

Theorem 6.9 Let $X_1, X_2, \ldots, X_n$ be an independent trials process with $E(X_j) = \mu$ and $V(X_j) = \sigma^2$. Let

$$S_n = X_1 + X_2 + \cdots + X_n$$

be the sum, and

$$A_n = \frac{S_n}{n}$$

be the average. Then

$$\begin{aligned}
E(S_n) &= n\mu\ , \\
V(S_n) &= n\sigma^2\ , \\
\sigma(S_n) &= \sigma\sqrt{n}\ , \\
E(A_n) &= \mu\ , \\
V(A_n) &= \frac{\sigma^2}{},\\
\sigma(A_n) &= \frac{\sigma}{\sqrt{n}}\ .
\end{aligned}$$

Proof. Since all the random variables X_j have the same expected value, we have

$$E(S_n) = E(X_1) + \cdots + E(X_n) = n\mu\ ,$$

$$V(S_n) = V(X_1) + \cdots + V(X_n) = n\sigma^2\ ,$$

and

$$\sigma(S_n) = \sigma\sqrt{n}\ .$$

We have seen that, if we multiply a random variable X with mean μ and variance σ^2 by a constant c, the new random variable has expected value $c\mu$ and variance $c^2\sigma^2$. Thus,

$$E(A_n) = E\left(\frac{S_n}{n}\right) = \frac{n\mu}{n} = \mu\ ,$$

and

$$V(A_n) = V\left(\frac{S_n}{n}\right) = \frac{V(S_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}\ .$$

Finally, the standard deviation of A_n is given by

$$\sigma(A_n) = \frac{\sigma}{\sqrt{n}}\ .$$

□

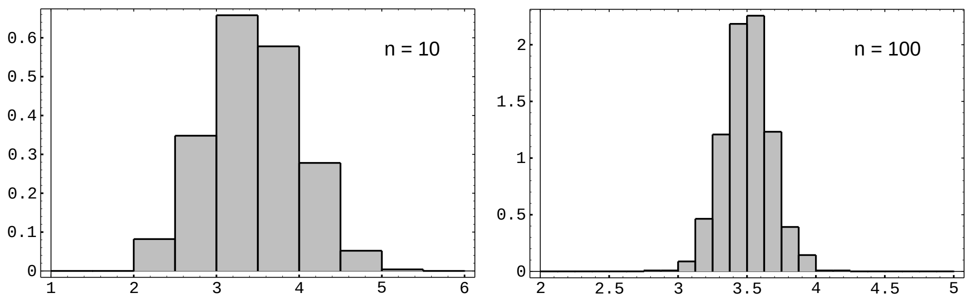

Figure 6.7: Empirical distribution of A_n.

The last equation in the above theorem implies that in an independent trials process, if the individual summands have finite variance, then the standard deviation of the average goes to 0 as $n \to \infty$. Since the standard deviation tells us something about the spread of the distribution around the mean, we see that for large values of n, the value of A_n is usually very close to the mean of A_n, which equals μ, as shown above. This statement is made precise in Chapter 8, where it is called the Law of Large Numbers. For example, let X represent the roll of a fair die. In Figure 6.7, we show the distribution of a random variable A_n corresponding to X, for $n = 10$ and $n = 100$.

Example 6.18 Consider n rolls of a die. We have seen that, if X_j is the outcome if the jth roll, then $E(X_j) = 7/2$ and $V(X_j) = 35/12$. Thus, if S_n is the sum of the outcomes, and $A_n = S_n/n$ is the average of the outcomes, we have $E(A_n) = 7/2$ and $V(A_n) = (35/12)/n$. Therefore, as n increases, the expected value of the average remains constant, but the variance tends to 0. If the variance is a measure of the expected deviation from the mean this would indicate that, for large n, we can expect the average to be very near the expected value. This is in fact the case, and we shall justify it in Chapter 8. $\square$

Bernoulli Trials

Consider next the general Bernoulli trials process. As usual, we let $X_j = 1$ if the jth outcome is a success and 0 if it is a failure. If p is the probability of a success, and $q = 1 - p$, then

$$\begin{aligned} E(X_j) &= 0q + 1p = p \ , \\ E(X_j^2) &= 0^2q + 1^2p = p \ , \end{aligned}$$

and

$$V(X_j) = E(X_j^2) - (E(X_j))^2 = p - p^2 = pq \ .$$

Thus, for Bernoulli trials, if $S_n = X_1 + X_2 + \cdots + X_n$ is the number of successes, then $E(S_n) = np$, $V(S_n) = npq$, and $D(S_n) = \sqrt{npq}$. If $A_n = S_n/n$ is the average number of successes, then $E(A_n) = p$, $V(A_n) = pq/n$, and $D(A_n) = \sqrt{pq/n}$. We see that the expected proportion of successes remains p and the variance tends to 0.

This suggests that the frequency interpretation of probability is a correct one. We shall make this more precise in Chapter 8.

Example 6.19 Let T denote the number of trials until the first success in a Bernoulli trials process. Then T is geometrically distributed. What is the variance of T? In Example 4.15, we saw that

$$m_T = \begin{pmatrix} 1 & 2 & 3 & \cdots \\ p & qp & q^2p & \cdots \end{pmatrix} .$$

In Example 6.4, we showed that

$$E(T) = 1/p \ .$$

Thus,

$$V(T) = E(T^2) - 1/p^2 \ ,$$

so we need only find

$$\begin{aligned} E(T^2) &= 1p + 4qp + 9q^2p + \cdots \\ &= p(1 + 4q + 9q^2 + \cdots) \ . \end{aligned}$$

To evaluate this sum, we start again with

$$1 + x + x^2 + \cdots = \frac{1}{1-x} \ .$$

Differentiating, we obtain

$$1 + 2x + 3x^2 + \cdots = \frac{1}{(1-x)^2} \ .$$

Multiplying by x,

$$x + 2x^2 + 3x^3 + \cdots = \frac{x}{(1-x)^2} \ .$$

Differentiating again gives

$$1 + 4x + 9x^2 + \cdots = \frac{1+x}{(1-x)^3} \ .$$

Thus,

$$E(T^2) = p\frac{1+q}{(1-q)^3} = \frac{1+q}{p^2}$$

and

$$\begin{aligned} V(T) &= E(T^2) - (E(T))^2 \\ &= \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} \ . \end{aligned}$$

For example, the variance for the number of tosses of a coin until the first head turns up is $(1/2)/(1/2)^2 = 2$. The variance for the number of rolls of a die until the first six turns up is $(5/6)/(1/6)^2 = 30$. Note that, as p decreases, the variance increases rapidly. This corresponds to the increased spread of the geometric distribution as p decreases (noted in Figure 5.1). □

Poisson Distribution

Just as in the case of expected values, it is easy to guess the variance of the Poisson distribution with parameter λ. We recall that the variance of a binomial distribution with parameters n and p equals npq. We also recall that the Poisson distribution could be obtained as a limit of binomial distributions, if n goes to ∞ and p goes to 0 in such a way that their product is kept fixed at the value λ. In this case, $npq = \lambda q$ approaches λ, since q goes to 1. So, given a Poisson distribution with parameter λ, we should guess that its variance is λ. The reader is asked to show this in Exercise 29.

Exercises

1 A number is chosen at random from the set $S = \{-1, 0, 1\}$. Let X be the number chosen. Find the expected value, variance, and standard deviation of X.

2 A random variable X has the distribution

$$p_X = \begin{pmatrix} 0 & 1 & 2 & 4 \\ 1/3 & 1/3 & 1/6 & 1/6 \end{pmatrix} .$$

Find the expected value, variance, and standard deviation of X.

3 You place a 1-dollar bet on the number 17 at Las Vegas, and your friend places a 1-dollar bet on black (see Exercises 1.1.6 and 1.1.7). Let X be your winnings and Y be her winnings. Compare $E(X)$, $E(Y)$, and $V(X)$, $V(Y)$. What do these computations tell you about the nature of your winnings if you and your friend make a sequence of bets, with you betting each time on a number and your friend betting on a color?

4 X is a random variable with $E(X) = 100$ and $V(X) = 15$. Find

(a) $E(X^2)$.

(b) $E(3X + 10)$.

(c) $E(-X)$.

(d) $V(-X)$.

(e) $D(-X)$.

5 In a certain manufacturing process, the (Fahrenheit) temperature never varies by more than 2° from 62°. The temperature is, in fact, a random variable F with distribution

$$P_F = \begin{pmatrix} 60 & 61 & 62 & 63 & 64 \\ 1/10 & 2/10 & 4/10 & 2/10 & 1/10 \end{pmatrix} .$$

(a) Find $E(F)$ and $V(F)$.

(b) Define $T = F - 62$. Find $E(T)$ and $V(T)$, and compare these answers with those in part (a).

(c) It is decided to report the temperature readings on a Celsius scale, that is, $C = (5/9)(F - 32)$. What is the expected value and variance for the readings now?

6 Write a computer program to calculate the mean and variance of a distribution which you specify as data. Use the program to compare the variances for the following densities, both having expected value 0:

$$p_X = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 3/11 & 2/11 & 1/11 & 2/11 & 3/11 \end{pmatrix} ;$$

$$p_Y = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 1/11 & 2/11 & 5/11 & 2/11 & 1/11 \end{pmatrix} .$$

7 A coin is tossed three times. Let X be the number of heads that turn up. Find $V(X)$ and $D(X)$.

8 A random sample of 2400 people are asked if they favor a government proposal to develop new nuclear power plants. If 40 percent of the people in the country are in favor of this proposal, find the expected value and the standard deviation for the number S_{2400} of people in the sample who favored the proposal.

9 A die is loaded so that the probability of a face coming up is proportional to the number on that face. The die is rolled with outcome X. Find $V(X)$ and $D(X)$.

10 Prove the following facts about the standard deviation.

(a) $D(X + c) = D(X)$.

(b) $D(cX) = |c|D(X)$.

11 A number is chosen at random from the integers 1, 2, 3, ..., n. Let X be the number chosen. Show that $E(X) = (n+1)/2$ and $V(X) = (n-1)(n+1)/12$. *Hint*: The following identity may be useful:

$$1^2 + 2^2 + \cdots + n^2 = \frac{(n)(n+1)(2n+1)}{6} .$$

12 Let X be a random variable with $\mu = E(X)$ and $\sigma^2 = V(X)$. Define $X^* = (X - \mu)/\sigma$. The random variable X^* is called the *standardized random variable* associated with X. Show that this standardized random variable has expected value 0 and variance 1.

13 Peter and Paul play Heads or Tails (see Example 1.4). Let W_n be Peter's winnings after n matches. Show that $E(W_n) = 0$ and $V(W_n) = n$.

14 Find the expected value and the variance for the number of boys and the number of girls in a royal family that has children until there is a boy or until there are three children, whichever comes first.

15 Suppose that n people have their hats returned at random. Let $X_i = 1$ if the ith person gets his or her own hat back and 0 otherwise. Let $S_n = \sum_{i=1}^n X_i$. Then S_n is the total number of people who get their own hats back. Show that

(a) $E(X_i^2) = 1/n$.

(b) $E(X_i \cdot X_j) = 1/n(n-1)$ for $i \neq j$.

(c) $E(S_n^2) = 2$ (using (a) and (b)).

(d) $V(S_n) = 1$.

16 Let S_n be the number of successes in n independent trials. Use the program **BinomialProbabilities** (Section 3.2) to compute, for given n, p, and j, the probability

$$P(-j\sqrt{npq} < S_n - np < j\sqrt{npq}) \ .$$

(a) Let $p = .5$, and compute this probability for $j = 1, 2, 3$ and $n = 10, 30, 50$. Do the same for $p = .2$.

(b) Show that the *standardized random variable* $S_n^* = (S_n - np)/\sqrt{npq}$ has expected value 0 and variance 1. What do your results from (a) tell you about this standardized quantity S_n^*?

17 Let X be the outcome of a chance experiment with $E(X) = \mu$ and $V(X) = \sigma^2$. When μ and σ^2 are unknown, the statistician often estimates them by repeating the experiment n times with outcomes x_1, x_2, ..., x_n, estimating μ by the *sample mean*

$$\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i \ ,$$

and σ^2 by the *sample variance*

$$s^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 \ .$$

Then s is the *sample standard deviation.* These formulas should remind the reader of the definitions of the theoretical mean and variance. (Many statisticians define the sample variance with the coefficient $1/n$ replaced by $1/(n-1)$. If this alternative definition is used, the expected value of s^2 is equal to σ^2. See Exercise 18, part (d).)

Write a computer program that will roll a die n times and compute the sample mean and sample variance. Repeat this experiment several times for $n = 10$ and $n = 1000$. How well do the sample mean and sample variance estimate the true mean 7/2 and variance 35/12?

18 Show that, for the sample mean $\bar{x}$ and sample variance s^2 as defined in Exercise 17,

(a) $E(\bar{x}) = \mu$.

(b) $E\big((\bar{x}-\mu)^2\big) = \sigma^2/n$.

(c) $E(s^2) = \frac{n-1}{n}\sigma^2$. *Hint*: For (c) write

$$\begin{aligned}
\sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n \big((x_i-\mu) - (\bar{x}-\mu)\big)^2 \\
&= \sum_{i=1}^n (x_i-\mu)^2 - 2(\bar{x}-\mu)\sum_{i=1}^n (x_i-\mu) + n(\bar{x}-\mu)^2 \\
&= \sum_{i=1}^n (x_i-\mu)^2 - n(\bar{x}-\mu)^2,
\end{aligned}$$

and take expectations of both sides, using part (b) when necessary.

(d) Show that if, in the definition of s^2 in Exercise 17, we replace the coefficient $1/n$ by the coefficient $1/(n-1)$, then $E(s^2) = \sigma^2$. (This shows why many statisticians use the coefficient $1/(n-1)$. The number s^2 is used to estimate the unknown quantity σ^2. If an estimator has an average value which equals the quantity being estimated, then the estimator is said to be *unbiased*. Thus, the statement $E(s^2) = \sigma^2$ says that s^2 is an unbiased estimator of σ^2.)

19 Let X be a random variable taking on values $a_1, a_2, \ldots, a_r$ with probabilities $p_1, p_2, \ldots, p_r$ and with $E(X) = \mu$. Define the *spread* of X as follows:

$$\bar{\sigma} = \sum_{i=1}^r |a_i - \mu| p_i \ .$$

This, like the standard deviation, is a way to quantify the amount that a random variable is spread out around its mean. Recall that the variance of a sum of mutually independent random variables is the sum of the individual variances. The square of the spread corresponds to the variance in a manner similar to the correspondence between the spread and the standard deviation. Show by an example that it is not necessarily true that the square of the spread of the sum of two independent random variables is the sum of the squares of the individual spreads.

20 We have two instruments that measure the distance between two points. The measurements given by the two instruments are random variables X_1 and X_2 that are independent with $E(X_1) = E(X_2) = \mu$, where μ is the true distance. From experience with these instruments, we know the values of the variances σ_1^2 and σ_2^2. These variances are not necessarily the same. From two measurements, we estimate μ by the weighted average $\bar{\mu} = wX_1 + (1-w)X_2$. Here w is chosen in $[0,1]$ to minimize the variance of $\bar{\mu}$.

(a) What is $E(\bar{\mu})$?

(b) How should w be chosen in $[0,1]$ to minimize the variance of $\bar{\mu}$?

21 Let X be a random variable with $E(X) = \mu$ and $V(X) = \sigma^2$. Show that the function $f(x)$ defined by

$$f(x) = \sum_\omega (X(\omega) - x)^2 p(\omega)$$

has its minimum value when $x = \mu$.

22 Let X and Y be two random variables defined on the finite sample space Ω. Assume that X, Y, $X + Y$, and $X - Y$ all have the same distribution. Prove that $P(X = Y = 0) = 1$.

23 If X and Y are any two random variables, then the *covariance* of X and Y is defined by $\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y)))$. Note that $\mathrm{Cov}(X, X) = V(X)$. Show that, if X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$; and show, by an example, that we can have $\mathrm{Cov}(X, Y) = 0$ and X and Y not independent.

***24** A professor wishes to make up a true-false exam with n questions. She assumes that she can design the problems in such a way that a student will answer the jth problem correctly with probability p_j, and that the answers to the various problems may be considered independent experiments. Let S_n be the number of problems that a student will get correct. The professor wishes to choose p_j so that $E(S_n) = .7n$ and so that the variance of S_n is as large as possible. Show that, to achieve this, she should choose $p_j = .7$ for all j; that is, she should make all the problems have the same difficulty.

25 (Lamperti[20]) An urn contains exactly 5000 balls, of which an unknown number X are white and the rest red, where X is a random variable with a probability distribution on the integers 0, 1, 2, . . . , 5000.

(a) Suppose we know that $E(X) = \mu$. Show that this is enough to allow us to calculate the probability that a ball drawn at random from the urn will be white. What is this probability?

(b) We draw a ball from the urn, examine its color, replace it, and then draw another. Under what conditions, if any, are the results of the two drawings independent; that is, does

$$P(\mathrm{white}, \mathrm{white}) = P(\mathrm{white})^2 \ ?$$

(c) Suppose the variance of X is σ^2. What is the probability of drawing two white balls in part (b)?

26 For a sequence of Bernoulli trials, let X_1 be the number of trials until the first success. For $j \geq 2$, let X_j be the number of trials after the $(j - 1)$st success until the jth success. It can be shown that X_1, X_2, . . . is an independent trials process.

[20] Private communication.

(a) What is the common distribution, expected value, and variance for X_j?

(b) Let $T_n = X_1 + X_2 + \cdots + X_n$. Then T_n is the time until the nth success. Find $E(T_n)$ and $V(T_n)$.

(c) Use the results of (b) to find the expected value and variance for the number of tosses of a coin until the nth occurrence of a head.

27 Referring to Exercise 6.1.30, find the variance for the number of boxes of Wheaties bought before getting half of the players' pictures and the variance for the number of additional boxes needed to get the second half of the players' pictures.

28 In Example 5.3, assume that the book in question has 1000 pages. Let X be the number of pages with no mistakes. Show that $E(X) = 905$ and $V(X) = 86$. Using these results, show that the probability is $\leq .05$ that there will be more than 924 pages without errors or fewer than 866 pages without errors.

29 Let X be Poisson distributed with parameter λ. Show that $V(X) = \lambda$.

6.3 Continuous Random Variables

In this section we consider the properties of the expected value and the variance of a continuous random variable. These quantities are defined just as for discrete random variables and share the same properties.

Expected Value

Definition 6.4 Let X be a real-valued random variable with density function $f(x)$. The *expected value* $\mu = E(X)$ is defined by

$$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx \ ,$$

provided the integral

$$\int_{-\infty}^{+\infty} |x| f(x)\, dx$$

is finite. □

The reader should compare this definition with the corresponding one for discrete random variables in Section 6.1. Intuitively, we can interpret $E(X)$, as we did in the previous sections, as the value that we should expect to obtain if we perform a large number of independent experiments and average the resulting values of X.

We can summarize the properties of $E(X)$ as follows (cf. Theorem 6.2).

Theorem 6.10 If X and Y are real-valued random variables and c is any constant, then

$$\begin{aligned} E(X+Y) &= E(X)+E(Y)\ , \\ E(cX) &= cE(X)\ . \end{aligned}$$

The proof is very similar to the proof of Theorem 6.2, and we omit it. □

More generally, if X_1, X_2, …, X_n are n real-valued random variables, and c_1, c_2, …, c_n are n constants, then

$$E(c_1X_1 + c_2X_2 + \cdots + c_nX_n) = c_1E(X_1) + c_2E(X_2) + \cdots + c_nE(X_n)\ .$$

Example 6.20 Let X be uniformly distributed on the interval $[0,1]$. Then

$$E(X) = \int_0^1 x\,dx = 1/2\ .$$

It follows that if we choose a large number N of random numbers from $[0,1]$ and take the average, then we can expect that this average should be close to the expected value of $1/2$. □

Example 6.21 Let $Z = (x,y)$ denote a point chosen uniformly and randomly from the unit disk, as in the dart game in Example 2.8 and let $X = (x^2+y^2)^{1/2}$ be the distance from Z to the center of the disk. The density function of X can easily be shown to equal $f(x) = 2x$, so by the definition of expected value,

$$\begin{aligned} E(X) &= \int_0^1 xf(x)\,dx \\ &= \int_0^1 x(2x)\,dx \\ &= \frac{2}{3}\ . \end{aligned}$$

□

Example 6.22 In the example of the couple meeting at the Inn (Example 2.16), each person arrives at a time which is uniformly distributed between 5:00 and 6:00 PM. The random variable Z under consideration is the length of time the first person has to wait until the second one arrives. It was shown that

$$f_Z(z) = 2(1-z)\ ,$$

for $0 \le z \le 1$. Hence,

$$E(Z) = \int_0^1 zf_Z(z)\,dz$$

$$
\begin{aligned}
&= \int_0^1 2z(1-z)\,dz \\
&= \left[z^2 - \frac{2}{3}z^3\right]_0^1 \\
&= \frac{1}{3}\ .
\end{aligned}
$$

□

Expectation of a Function of a Random Variable

Suppose that X is a real-valued random variable and $\phi(x)$ is a continuous function from $\mathbf{R}$ to $\mathbf{R}$. The following theorem is the continuous analogue of Theorem 6.1.

Theorem 6.11 If X is a real-valued random variable and if $\phi : \mathbf{R} \to \mathbf{R}$ is a continuous real-valued function with domain $[a, b]$, then

$$
E(\phi(X)) = \int_{-\infty}^{+\infty} \phi(x) f_X(x)\,dx\ ,
$$

provided the integral exists. □

For a proof of this theorem, see Ross.[21]

Expectation of the Product of Two Random Variables

In general, it is not true that $E(XY) = E(X)E(Y)$, since the integral of a product is not the product of integrals. But if X and Y are independent, then the expectations multiply.

Theorem 6.12 Let X and Y be independent real-valued continuous random variables with finite expected values. Then we have

$$
E(XY) = E(X)E(Y)\ .
$$

Proof. We will prove this only in the case that the ranges of X and Y are contained in the intervals $[a, b]$ and $[c, d]$, respectively. Let the density functions of X and Y be denoted by $f_X(x)$ and $f_Y(y)$, respectively. Since X and Y are independent, the joint density function of X and Y is the product of the individual density functions. Hence

$$
\begin{aligned}
E(XY) &= \int_a^b \int_c^d xy f_X(x) f_Y(y)\,dy\,dx \\
&= \int_a^b x f_X(x)\,dx \int_c^d y f_Y(y)\,dy \\
&= E(X)E(Y)\ .
\end{aligned}
$$

The proof in the general case involves using sequences of bounded random variables that approach X and Y, and is somewhat technical, so we will omit it. □

[21] S. Ross, *A First Course in Probability,* (New York: Macmillan, 1984), pgs. 241-245.

In the same way, one can show that if X_1, X_2, …, X_n are n mutually independent real-valued random variables, then

$$E(X_1X_2\cdots X_n) = E(X_1)\,E(X_2)\,\cdots\,E(X_n)\ .$$

Example 6.23 Let $Z = (X, Y)$ be a point chosen at random in the unit square. Let $A = X^2$ and $B = Y^2$. Then Theorem 4.3 implies that A and B are independent. Using Theorem 6.11, the expectations of A and B are easy to calculate:

$$\begin{aligned} E(A) = E(B) &= \int_0^1 x^2\,dx \\ &= \frac{1}{3}\ . \end{aligned}$$

Using Theorem 6.12, the expectation of AB is just the product of $E(A)$ and $E(B)$, or 1/9. The usefulness of this theorem is demonstrated by noting that it is quite a bit more difficult to calculate $E(AB)$ from the definition of expectation. One finds that the density function of AB is

$$f_{AB}(t) = \frac{-\log(t)}{4\sqrt{t}}\ ,$$

so

$$\begin{aligned} E(AB) &= \int_0^1 t f_{AB}(t)\,dt \\ &= \frac{1}{9}\ . \end{aligned}$$

□

Example 6.24 Again let $Z = (X, Y)$ be a point chosen at random in the unit square, and let $W = X + Y$. Then Y and W are not independent, and we have

$$\begin{aligned} E(Y) &= \frac{1}{2}\ , \\ E(W) &= 1\ , \\ E(YW) &= E(XY + Y^2) = E(X)E(Y) + \frac{1}{3} = \frac{7}{12} \neq E(Y)E(W)\ . \end{aligned}$$

□

We turn now to the variance.

Variance

Definition 6.5 Let X be a real-valued random variable with density function $f(x)$. The *variance* $\sigma^2 = V(X)$ is defined by

$$\sigma^2 = V(X) = E((X - \mu)^2)\ .$$

□

The next result follows easily from Theorem 6.1. There is another way to calculate the variance of a continuous random variable, which is usually slightly easier. It is given in Theorem 6.15.

Theorem 6.13 If X is a real-valued random variable with $E(X) = \mu$, then

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx \ .$$

□

The properties listed in the next three theorems are all proved in exactly the same way that the corresponding theorems for discrete random variables were proved in Section 6.2.

Theorem 6.14 If X is a real-valued random variable defined on Ω and c is any constant, then (cf. Theorem 6.7)

$$\begin{aligned} V(cX) &= c^2 V(X) \ , \\ V(X + c) &= V(X) \ . \end{aligned}$$

□

Theorem 6.15 If X is a real-valued random variable with $E(X) = \mu$, then (cf. Theorem 6.6)

$$V(X) = E(X^2) - \mu^2 \ .$$

□

Theorem 6.16 If X and Y are independent real-valued random variables on Ω, then (cf. Theorem 6.8)

$$V(X + Y) = V(X) + V(Y) \ .$$

□

Example 6.25 (continuation of Example 6.20) If X is uniformly distributed on $[0, 1]$, then, using Theorem 6.15, we have

$$V(X) = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \frac{1}{12} \ .$$

□

Example 6.26 Let X be an exponentially distributed random variable with parameter λ. Then the density function of X is

$$f_X(x) = \lambda e^{-\lambda x} \ .$$

From the definition of expectation and integration by parts, we have

$$\begin{aligned} E(X) &= \int_0^\infty x f_X(x)\,dx \\ &= \lambda \int_0^\infty x e^{-\lambda x}\,dx \\ &= -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\,dx \\ &= 0 + \frac{e^{-\lambda x}}{-\lambda}\Big|_0^\infty = \frac{1}{\lambda} \ . \end{aligned}$$

Similarly, using Theorems 6.11 and 6.15, we have

$$\begin{aligned} V(X) &= \int_0^\infty x^2 f_X(x)\,dx - \frac{1}{\lambda^2} \\ &= \lambda \int_0^\infty x^2 e^{-\lambda x}\,dx - \frac{1}{\lambda^2} \\ &= -x^2 e^{-\lambda x}\Big|_0^\infty + 2\int_0^\infty x e^{-\lambda x}\,dx - \frac{1}{\lambda^2} \\ &= -x^2 e^{-\lambda x}\Big|_0^\infty - \frac{2xe^{-\lambda x}}{\lambda}\Big|_0^\infty - \frac{2}{\lambda^2} e^{-\lambda x}\Big|_0^\infty - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \ . \end{aligned}$$

In this case, both $E(X)$ and $V(X)$ are finite if $\lambda > 0$. $\square$

Example 6.27 Let Z be a standard normal random variable with density function

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ .$$

Since this density function is symmetric with respect to the y-axis, then it is easy to show that

$$\int_{-\infty}^{\infty} x f_Z(x)\,dx$$

has value 0. The reader should recall however, that the expectation is defined to be the above integral only if the integral

$$\int_{-\infty}^{\infty} |x| f_Z(x)\,dx$$

is finite. This integral equals

$$2\int_0^\infty x f_Z(x)\,dx \ ,$$

which one can easily show is finite. Thus, the expected value of Z is 0.

To calculate the variance of Z, we begin by applying Theorem 6.15:

$$V(Z) = \int_{-\infty}^{+\infty} x^2 f_Z(x)\, dx - \mu^2 \ .$$

If we write x^2 as $x \cdot x$, and integrate by parts, we obtain

$$\frac{1}{\sqrt{2\pi}}(-xe^{-x^2/2})\Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2}\, dx \ .$$

The first summand above can be shown to equal 0, since as $x \to \pm\infty$, $e^{-x^2/2}$ gets small more quickly than x gets large. The second summand is just the standard normal density integrated over its domain, so the value of this summand is 1. Therefore, the variance of the standard normal density equals 1.

Now let X be a (not necessarily standard) normal random variable with parameters μ and σ. Then the density function of X is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \ .$$

We can write $X = \sigma Z + \mu$, where Z is a standard normal random variable. Since $E(Z) = 0$ and $V(Z) = 1$ by the calculation above, Theorems 6.10 and 6.14 imply that

$$\begin{aligned} E(X) &= E(\sigma Z + \mu) = \mu \ , \\ V(X) &= V(\sigma Z + \mu) = \sigma^2 \ . \end{aligned}$$

□

Example 6.28 Let X be a continuous random variable with the Cauchy density function

$$f_X(x) = \frac{a}{\pi} \frac{1}{a^2 + x^2} \ .$$

Then the expectation of X does not exist, because the integral

$$\frac{a}{\pi} \int_{-\infty}^{+\infty} \frac{|x|\, dx}{a^2 + x^2}$$

diverges. Thus the variance of X also fails to exist. Densities whose variance is not defined, like the Cauchy density, behave quite differently in a number of important respects from those whose variance is finite. We shall see one instance of this difference in Section 8.2. □

Independent Trials

Corollary 6.1 If $X_1, X_2, \ldots, X_n$ is an independent trials process of real-valued random variables, with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$, and if

$$\begin{aligned} S_n &= X_1 + X_2 + \cdots + X_n \ , \\ A_n &= \frac{S_n}{n} \ , \end{aligned}$$

then

$$\begin{aligned} E(S_n) &= n\mu \ , \\ E(A_n) &= \mu \ , \\ V(S_n) &= n\sigma^2 \ , \\ V(A_n) &= \frac{\sigma^2}{n} \ . \end{aligned}$$

It follows that if we set

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \ ,$$

then

$$\begin{aligned} E(S_n^*) &= 0 \ , \\ V(S_n^*) &= 1 \ . \end{aligned}$$

We say that S_n^* is a *standardized version of* S_n (see Exercise 12 in Section 6.2). □

Queues

Example 6.29 Let us consider again the queueing problem, that is, the problem of the customers waiting in a queue for service (see Example 5.7). We suppose again that customers join the queue in such a way that the time between arrivals is an exponentially distributed random variable X with density function

$$f_X(t) = \lambda e^{-\lambda t} \ .$$

Then the expected value of the time between arrivals is simply $1/\lambda$ (see Example 6.26), as was stated in Example 5.7. The reciprocal λ of this expected value is often referred to as the *arrival rate.* The *service time* of an individual who is first in line is defined to be the amount of time that the person stays at the head of the line before leaving. We suppose that the customers are served in such a way that the service time is another exponentially distributed random variable Y with density function

$$f_X(t) = \mu e^{-\mu t} \ .$$

Then the expected value of the service time is

$$E(X) = \int_0^\infty t f_X(t)\, dt = \frac{1}{\mu} \ .$$

The reciprocal μ if this expected value is often referred to as the *service rate.*

We expect on grounds of our everyday experience with queues that if the service rate is greater than the arrival rate, then the average queue size will tend to stabilize, but if the service rate is less than the arrival rate, then the queue will tend to increase in length without limit (see Figure 5.7). The simulations in Example 5.7 tend to bear out our everyday experience. We can make this conclusion more precise if we introduce the *traffic intensity* as the product

$$\rho = (\text{arrival rate})(\text{average service time}) = \frac{\lambda}{\mu} = \frac{1/\mu}{1/\lambda} \ .$$

The traffic intensity is also the ratio of the average service time to the average time between arrivals. If the traffic intensity is less than 1 the queue will perform reasonably, but if it is greater than 1 the queue will grow indefinitely large. In the critical case of $\rho = 1$, it can be shown that the queue will become large but there will always be times at which the queue is empty.[22]

In the case that the traffic intensity is less than 1 we can consider the length of the queue as a random variable Z whose expected value is finite,

$$E(Z) = N \ .$$

The time spent in the queue by a single customer can be considered as a random variable W whose expected value is finite,

$$E(W) = T \ .$$

Then we can argue that, when a customer joins the queue, he expects to find N people ahead of him, and when he leaves the queue, he expects to find λT people behind him. Since, in equilibrium, these should be the same, we would expect to find that

$$N = \lambda T \ .$$

This last relationship is called *Little's law for queues.*[23] We will not prove it here. A proof may be found in Ross.[24] Note that in this case we are counting the waiting time of all customers, even those that do not have to wait at all. In our simulation in Section 4.2, we did not consider these customers.

If we knew the expected queue length then we could use Little's law to obtain the expected waiting time, since

$$T = \frac{N}{\lambda} \ .$$

The queue length is a random variable with a discrete distribution. We can estimate this distribution by simulation, keeping track of the queue lengths at the times at which a customer arrives. We show the result of this simulation (using the program **Queue**) in Figure 6.8.

[22]L. Kleinrock, *Queueing Systems,* vol. 2 (New York: John Wiley and Sons, 1975).
[23]ibid., p. 17.
[24]S. M. Ross, *Applied Probability Models with Optimization Applications,* (San Francisco: Holden-Day, 1970)

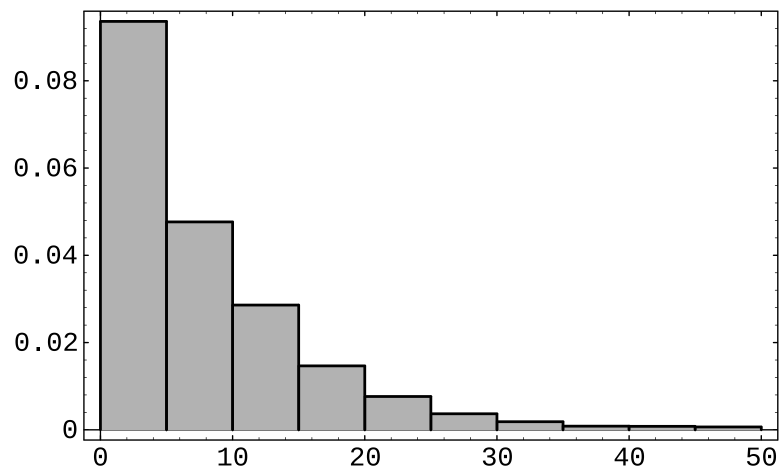

Figure 6.8: Distribution of queue lengths.

We note that the distribution appears to be a geometric distribution. In the study of queueing theory it is shown that the distribution for the queue length in equilibrium is indeed a geometric distribution with

$$s_j = (1-\rho)\rho^j \qquad \text{for } j = 0, 1, 2, \dots\ ,$$

if $\rho < 1$. The expected value of a random variable with this distribution is

$$N = \frac{\rho}{(1-\rho)}$$

(see Example 6.4). Thus by Little's result the expected waiting time is

$$T = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu - \lambda}\ ,$$

where μ is the service rate, λ the arrival rate, and ρ the traffic intensity.

In our simulation, the arrival rate is 1 and the service rate is 1.1. Thus, the traffic intensity is $1/1.1 = 10/11$, the expected queue size is

$$\frac{10/11}{(1-10/11)} = 10\ ,$$

and the expected waiting time is

$$\frac{1}{1.1-1} = 10\ .$$

In our simulation the average queue size was 8.19 and the average waiting time was 7.37. In Figure 6.9, we show the histogram for the waiting times. This histogram suggests that the density for the waiting times is exponential with parameter $\mu - \lambda$, and this is the case. $\square$

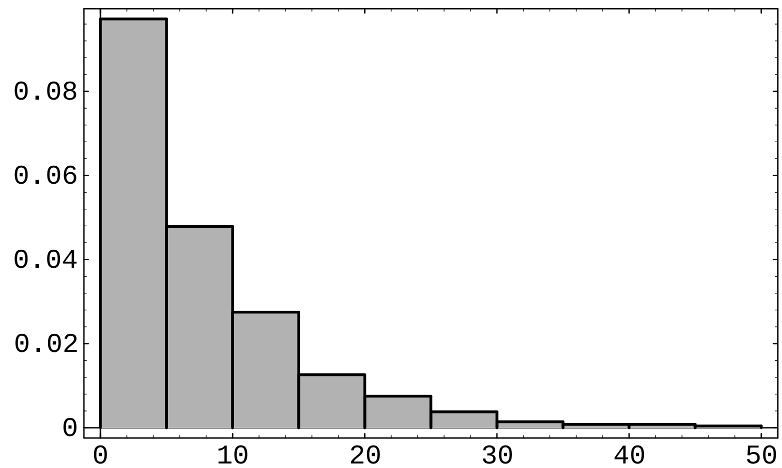

Figure 6.9: Distribution of queue waiting times.

Exercises

1 Let X be a random variable with range $[-1, 1]$ and let $f_X(x)$ be the density function of X. Find $\mu(X)$ and $\sigma^2(X)$ if, for $|x| < 1$,

(a) $f_X(x) = 1/2$.

(b) $f_X(x) = |x|$.

(c) $f_X(x) = 1 - |x|$.

(d) $f_X(x) = (3/2)x^2$.

2 Let X be a random variable with range $[-1, 1]$ and f_X its density function. Find $\mu(X)$ and $\sigma^2(X)$ if, for $|x| > 1$, $f_X(x) = 0$, and for $|x| < 1$,

(a) $f_X(x) = (3/4)(1 - x^2)$.

(b) $f_X(x) = (\pi/4)\cos(\pi x/2)$.

(c) $f_X(x) = (x + 1)/2$.

(d) $f_X(x) = (3/8)(x + 1)^2$.

3 The lifetime, measure in hours, of the ACME super light bulb is a random variable T with density function $f_T(t) = \lambda^2 t e^{-\lambda t}$, where $\lambda = .05$. What is the expected lifetime of this light bulb? What is its variance?

4 Let X be a random variable with range $[-1, 1]$ and density function $f_X(x) = ax + b$ if $|x| < 1$.

(a) Show that if $\int_{-1}^{+1} f_X(x)\,dx = 1$, then $b = 1/2$.

(b) Show that if $f_X(x) \geq 0$, then $-1/2 \leq a \leq 1/2$.

(c) Show that $\mu = (2/3)a$, and hence that $-1/3 \leq \mu \leq 1/3$.

(d) Show that $\sigma^2(X) = (2/3)b - (4/9)a^2 = 1/3 - (4/9)a^2$.

5 Let X be a random variable with range $[-1, 1]$ and density function $f_X(x) = ax^2 + bx + c$ if $|x| < 1$ and 0 otherwise.

(a) Show that $2a/3 + 2c = 1$ (see Exercise 4).

(b) Show that $2b/3 = \mu(X)$.

(c) Show that $2a/5 + 2c/3 = \sigma^2(X)$.

(d) Find a, b, and c if $\mu(X) = 0$, $\sigma^2(X) = 1/15$, and sketch the graph of f_X.

(e) Find a, b, and c if $\mu(X) = 0$, $\sigma^2(X) = 1/2$, and sketch the graph of f_X.

6 Let T be a random variable with range $[0, \infty]$ and f_T its density function. Find $\mu(T)$ and $\sigma^2(T)$ if, for $t < 0$, $f_T(t) = 0$, and for $t > 0$,

(a) $f_T(t) = 3e^{-3t}$.

(b) $f_T(t) = 9te^{-3t}$.

(c) $f_T(t) = 3/(1 + t)^4$.

7 Let X be a random variable with density function f_X. Show, using elementary calculus, that the function

$$\phi(a) = E((X - a)^2)$$

takes its minimum value when $a = \mu(X)$, and in that case $\phi(a) = \sigma^2(X)$.

8 Let X be a random variable with mean μ and variance σ^2. Let $Y = aX^2 + bX + c$. Find the expected value of Y.

9 Let X, Y, and Z be independent random variables, each with mean μ and variance σ^2.

(a) Find the expected value and variance of $S = X + Y + Z$.

(b) Find the expected value and variance of $A = (1/3)(X + Y + Z)$.

(c) Find the expected value of S^2 and A^2.

10 Let X and Y be independent random variables with uniform density functions on $[0, 1]$. Find

(a) $E(|X - Y|)$.

(b) $E(\max(X, Y))$.

(c) $E(\min(X, Y))$.

(d) $E(X^2 + Y^2)$.

(e) $E((X + Y)^2)$.

11 The Pilsdorff Beer Company runs a fleet of trucks along the 100 mile road from Hangtown to Dry Gulch. The trucks are old, and are apt to break down at any point along the road with equal probability. Where should the company locate a garage so as to minimize the expected distance from a typical breakdown to the garage? In other words, if X is a random variable giving the location of the breakdown, measured, say, from Hangtown, and b gives the location of the garage, what choice of b minimizes $E(|X-b|)$? Now suppose X is not distributed uniformly over $[0,100]$, but instead has density function $f_X(x) = 2x/10{,}000$. Then what choice of b minimizes $E(|X-b|)$?

12 Find $E(X^Y)$, where X and Y are independent random variables which are uniform on $[0,1]$. Then verify your answer by simulation.

13 Let X be a random variable that takes on nonnegative values and has distribution function $F(x)$. Show that

$$E(X) = \int_0^\infty (1 - F(x))\, dx\ .$$

Hint: Integrate by parts.

Illustrate this result by calculating $E(X)$ by this method if X has an exponential distribution $F(x) = 1 - e^{-\lambda x}$ for $x \geq 0$, and $F(x) = 0$ otherwise.

14 Let X be a continuous random variable with density function $f_X(x)$. Show that if

$$\int_{-\infty}^{+\infty} x^2 f_X(x)\, dx < \infty\ ,$$

then

$$\int_{-\infty}^{+\infty} |x| f_X(x)\, dx < \infty\ .$$

Hint: Except on the interval $[-1,1]$, the first integrand is greater than the second integrand.

15 Let X be a random variable distributed uniformly over $[0,20]$. Define a new random variable Y by $Y = \lfloor X \rfloor$ (the greatest integer in X). Find the expected value of Y. Do the same for $Z = \lfloor X + .5 \rfloor$. Compute $E\bigl(|X-Y|\bigr)$ and $E\bigl(|X-Z|\bigr)$. (Note that Y is the value of X rounded off to the nearest smallest integer, while Z is the value of X rounded off to the nearest integer. Which method of rounding off is better? Why?)

16 Assume that the lifetime of a diesel engine part is a random variable X with density f_X. When the part wears out, it is replaced by another with the same density. Let $N(t)$ be the number of parts that are used in time t. We want to study the random variable $N(t)/t$. Since parts are replaced on the average every $E(X)$ time units, we expect about $t/E(X)$ parts to be used in time t. That is, we expect that

$$\lim_{t\to\infty} E\Bigl(\frac{N(t)}{t}\Bigr) = \frac{1}{E(X)}\ .$$

This result is correct but quite difficult to prove. Write a program that will allow you to specify the density f_X, and the time t, and simulate this experiment to find $N(t)/t$. Have your program repeat the experiment 500 times and plot a bar graph for the random outcomes of $N(t)/t$. From this data, estimate $E(N(t)/t)$ and compare this with $1/E(X)$. In particular, do this for $t = 100$ with the following two densities:

(a) $f_X = e^{-t}$.

(b) $f_X = te^{-t}$.

17 Let X and Y be random variables. The *covariance* Cov(X, Y) is defined by (see Exercise 6.2.23)

$$\mathrm{cov}(X, Y) = E((X - \mu(X))(Y - \mu(Y))) \ .$$

(a) Show that $\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$.

(b) Using (a), show that $\mathrm{cov}(X, Y) = 0$, if X and Y are independent. (Caution: the converse is *not* always true.)

(c) Show that $V(X + Y) = V(X) + V(Y) + 2\mathrm{cov}(X, Y)$.

18 Let X and Y be random variables with positive variance. The *correlation* of X and Y is defined as

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{V(X)V(Y)}} \ .$$

(a) Using Exercise 17(c), show that

$$0 \le V\left(\frac{X}{\sigma(X)} + \frac{Y}{\sigma(Y)}\right) = 2(1 + \rho(X, Y)) \ .$$

(b) Now show that

$$0 \le V\left(\frac{X}{\sigma(X)} - \frac{Y}{\sigma(Y)}\right) = 2(1 - \rho(X, Y)) \ .$$

(c) Using (a) and (b), show that

$$-1 \le \rho(X, Y) \le 1 \ .$$

19 Let X and Y be independent random variables with uniform densities in $[0, 1]$. Let $Z = X + Y$ and $W = X - Y$. Find

(a) $\rho(X, Y)$ (see Exercise 18).

(b) $\rho(X, Z)$.

(c) $\rho(Y, W)$.

(d) $\rho(Z, W)$.

***20** When studying certain physiological data, such as heights of fathers and sons, it is often natural to assume that these data (e.g., the heights of the fathers and the heights of the sons) are described by random variables with normal densities. These random variables, however, are not independent but rather are correlated. For example, a two-dimensional standard normal density for correlated random variables has the form

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot e^{-(x^2-2\rho xy+y^2)/2(1-\rho^2)} \ .$$

(a) Show that X and Y each have standard normal densities.

(b) Show that the correlation of X and Y (see Exercise 18) is ρ.

***21** For correlated random variables X and Y it is natural to ask for the expected value for X given Y. For example, Galton calculated the expected value of the height of a son given the height of the father. He used this to show that tall men can be expected to have sons who are less tall on the average. Similarly, students who do very well on one exam can be expected to do less well on the next exam, and so forth. This is called *regression on the mean.* To define this conditional expected value, we first define a conditional density of X given $Y = y$ by

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

where $f_{X,Y}(x,y)$ is the joint density of X and Y, and f_Y is the density for Y. Then the conditional expected value of X given Y is

$$E(X|Y=y) = \int_a^b x f_{X|Y}(x|y)\, dx \ .$$

For the normal density in Exercise 20, show that the conditional density of $f_{X|Y}(x|y)$ is normal with mean ρy and variance $1-\rho^2$. From this we see that if X and Y are positively correlated $(0 < \rho < 1)$, and if $y > E(Y)$, then the expected value for X given $Y = y$ will be less than y (i.e., we have regression on the mean).

22 A point Y is chosen at random from $[0,1]$. A second point X is then chosen from the interval $[0,Y]$. Find the density for X. *Hint*: Calculate $f_{X|Y}$ as in Exercise 21 and then use

$$f_X(x) = \int_x^1 f_{X|Y}(x|y) f_Y(y)\, dy \ .$$

Can you also derive your result geometrically?

***23** Let X and V be two standard normal random variables. Let ρ be a real number between -1 and 1.

(a) Let $Y = \rho X + \sqrt{1-\rho^2}V$. Show that $E(Y) = 0$ and $Var(Y) = 1$. We shall see later (see Example 7.5 and Example 10.17), that the sum of two independent normal random variables is again normal. Thus, assuming this fact, we have shown that Y is standard normal.

(b) Using Exercises 17 and 18, show that the correlation of X and Y is ρ.

(c) In Exercise 20, the joint density function $f_{X,Y}(x, y)$ for the random variable (X, Y) is given. Now suppose that we want to know the set of points (x, y) in the xy-plane such that $f_{X,Y}(x, y) = C$ for some constant C. This set of points is called a set of constant density. Roughly speaking, a set of constant density is a set of points where the outcomes (X, Y) are equally likely to fall. Show that for a given C, the set of points of constant density is a curve whose equation is

$$x^2 - 2\rho xy + y^2 = D ,$$

where D is a constant which depends upon C. (This curve is an ellipse.)

(d) One can plot the ellipse in part (c) by using the parametric equations

$$\begin{aligned} x &= \frac{r\cos\theta}{\sqrt{2(1-\rho)}} + \frac{r\sin\theta}{\sqrt{2(1+\rho)}} , \\ y &= \frac{r\cos\theta}{\sqrt{2(1-\rho)}} - \frac{r\sin\theta}{\sqrt{2(1+\rho)}} . \end{aligned}$$

Write a program to plot 1000 pairs (X, Y) for $\rho = -1/2, 0, 1/2$. For each plot, have your program plot the above parametric curves for $r = 1, 2, 3$.

***24** Following Galton, let us assume that the fathers and sons have heights that are dependent normal random variables. Assume that the average height is 68 inches, standard deviation is 2.7 inches, and the correlation coefficient is .5 (see Exercises 20 and 21). That is, assume that the heights of the fathers and sons have the form $2.7X + 68$ and $2.7Y + 68$, respectively, where X and Y are correlated standardized normal random variables, with correlation coefficient .5.

(a) What is the expected height for the son of a father whose height is 72 inches?

(b) Plot a scatter diagram of the heights of 1000 father and son pairs. *Hint*: You can choose standardized pairs as in Exercise 23 and then plot $(2.7X+ 68, 2.7Y + 68)$.

***25** When we have pairs of data (x_i, y_i) that are outcomes of the pairs of dependent random variables X, Y we can estimate the coorelation coefficient ρ by

$$\bar{r} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_X s_Y} ,$$

where $\bar{x}$ and $\bar{y}$ are the sample means for X and Y, respectively, and s_X and s_Y are the sample standard deviations for X and Y (see Exercise 6.2.17). Write a program to compute the sample means, variances, and correlation for such dependent data. Use your program to compute these quantities for Galton's data on heights of parents and children given in Appendix B.

Plot the equal density ellipses as defined in Exercise 23 for $r = 4$, 6, and 8, and on the same graph print the values that appear in the table at the appropriate points. For example, print 12 at the point $(70.5, 68.2)$, indicating that there were 12 cases where the parent's height was 70.5 and the child's was 68.12. See if Galton's data is consistent with the equal density ellipses.

26 (from Hamming[25]) Suppose you are standing on the bank of a straight river.

(a) Choose, at random, a direction which will keep you on dry land, and walk 1 km in that direction. Let P denote your position. What is the expected distance from P to the river?

(b) Now suppose you proceed as in part (a), but when you get to P, you pick a random direction (from among *all* directions) and walk 1 km. What is the probability that you will reach the river before the second walk is completed?

27 (from Hamming[26]) A game is played as follows: A random number X is chosen uniformly from $[0, 1]$. Then a sequence $Y_1, Y_2, \ldots$ of random numbers is chosen independently and uniformly from $[0, 1]$. The game ends the first time that $Y_i > X$. You are then paid $(i - 1)$ dollars. What is a fair entrance fee for this game?

28 A long needle of length L much bigger than 1 is dropped on a grid with horizontal and vertical lines one unit apart. Show that the average number a of lines crossed is approximately

$$a = \frac{4L}{\pi} \ .$$

[25] R. W. Hamming, *The Art of Probability for Scientists and Engineers* (Redwood City: Addison-Wesley, 1991), p. 192.
[26] ibid., pg. 205.

Chapter 7

Sums of Independent Random Variables

7.1 Sums of Discrete Random Variables

In this chapter we turn to the important question of determining the distribution of a sum of independent random variables in terms of the distributions of the individual constituents. In this section we consider only sums of discrete random variables, reserving the case of continuous random variables for the next section.

We consider here only random variables whose values are integers. Their distribution functions are then defined on these integers. We shall find it convenient to assume here that these distribution functions are defined for *all* integers, by defining them to be 0 where they are not otherwise defined.

Convolutions

Suppose X and Y are two independent discrete random variables with distribution functions $m_1(x)$ and $m_2(x)$. Let $Z = X + Y$. We would like to determine the distribution function $m_3(x)$ of Z. To do this, it is enough to determine the probability that Z takes on the value z, where z is an arbitrary integer. Suppose that $X = k$, where k is some integer. Then $Z = z$ if and only if $Y = z - k$. So the event $Z = z$ is the union of the pairwise disjoint events

$$(X = k) \text{ and } (Y = z - k) ,$$

where k runs over the integers. Since these events are pairwise disjoint, we have

$$P(Z = z) = \sum_{k=-\infty}^{\infty} P(X = k) \cdot P(Y = z - k) .$$

Thus, we have found the distribution function of the random variable Z. This leads to the following definition.

Definition 7.1 Let X and Y be two independent integer-valued random variables, with distribution functions $m_1(x)$ and $m_2(x)$ respectively. Then the *convolution* of $m_1(x)$ and $m_2(x)$ is the distribution function $m_3 = m_1 * m_2$ given by

$$m_3(j) = \sum_k m_1(k) \cdot m_2(j-k) \ ,$$

for $j = \ldots,\ -2,\ -1,\ 0,\ 1,\ 2,\ \ldots$. The function $m_3(x)$ is the distribution function of the random variable $Z = X + Y$. □

It is easy to see that the convolution operation is commutative, and it is straightforward to show that it is also associative.

Now let $S_n = X_1 + X_2 + \cdots + X_n$ be the sum of n independent random variables of an independent trials process with common distribution function m defined on the integers. Then the distribution function of S_1 is m. We can write

$$S_n = S_{n-1} + X_n \ .$$

Thus, since we know the distribution function of X_n is m, we can find the distribution function of S_n by induction.

Example 7.1 A die is rolled twice. Let X_1 and X_2 be the outcomes, and let $S_2 = X_1 + X_2$ be the sum of these outcomes. Then X_1 and X_2 have the common distribution function:

$$m = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

The distribution function of S_2 is then the convolution of this distribution with itself. Thus,

$$\begin{aligned} P(S_2 = 2) &= m(1)m(1) \\ &= \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \ , \\ P(S_2 = 3) &= m(1)m(2) + m(2)m(1) \\ &= \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{2}{36} \ , \\ P(S_2 = 4) &= m(1)m(3) + m(2)m(2) + m(3)m(1) \\ &= \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{3}{36} \ . \end{aligned}$$

Continuing in this way we would find $P(S_2 = 5) = 4/36$, $P(S_2 = 6) = 5/36$, $P(S_2 = 7) = 6/36$, $P(S_2 = 8) = 5/36$, $P(S_2 = 9) = 4/36$, $P(S_2 = 10) = 3/36$, $P(S_2 = 11) = 2/36$, and $P(S_2 = 12) = 1/36$.

The distribution for S_3 would then be the convolution of the distribution for S_2 with the distribution for X_3. Thus

$$P(S_3 = 3) = P(S_2 = 2)P(X_3 = 1)$$

$$
\begin{aligned}
&= \frac{1}{36}\cdot\frac{1}{6} = \frac{1}{216}\ , \\
P(S_3 = 4) &= P(S_2 = 3)P(X_3 = 1) + P(S_2 = 2)P(X_3 = 2) \\
&= \frac{2}{36}\cdot\frac{1}{6} + \frac{1}{36}\cdot\frac{1}{6} = \frac{3}{216}\ ,
\end{aligned}
$$

and so forth.

This is clearly a tedious job, and a program should be written to carry out this calculation. To do this we first write a program to form the convolution of two densities p and q and return the density r. We can then write a program to find the density for the sum S_n of n independent random variables with a common density p, at least in the case that the random variables have a finite number of possible values.

Running this program for the example of rolling a die n times for $n = 10,\ 20,\ 30$ results in the distributions shown in Figure 7.1. We see that, as in the case of Bernoulli trials, the distributions become bell-shaped. We shall discuss in Chapter 9 a very general theorem called the *Central Limit Theorem* that will explain this phenomenon. □

Example 7.2 A well-known method for evaluating a bridge hand is: an ace is assigned a value of 4, a king 3, a queen 2, and a jack 1. All other cards are assigned a value of 0. The *point count* of the hand is then the sum of the values of the cards in the hand. (It is actually more complicated than this, taking into account voids in suits, and so forth, but we consider here this simplified form of the point count.) If a card is dealt at random to a player, then the point count for this card has distribution

$$
p_X = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 36/52 & 4/52 & 4/52 & 4/52 & 4/52 \end{pmatrix}.
$$

Let us regard the total hand of 13 cards as 13 independent trials with this common distribution. (Again this is not quite correct because we assume here that we are always choosing a card from a full deck.) Then the distribution for the point count C for the hand can be found from the program **NFoldConvolution** by using the distribution for a single card and choosing $n = 13$. A player with a point count of 13 or more is said to have an *opening bid.* The probability of having an opening bid is then

$$
P(C \geq 13)\ .
$$

Since we have the distribution of C, it is easy to compute this probability. Doing this we find that

$$
P(C \geq 13) = .2845\ ,
$$

so that about one in four hands should be an opening bid according to this simplified model. A more realistic discussion of this problem can be found in Epstein, *The Theory of Gambling and Statistical Logic.*[1] □

[1] R. A. Epstein, *The Theory of Gambling and Statistical Logic,* rev. ed. (New York: Academic Press, 1977).

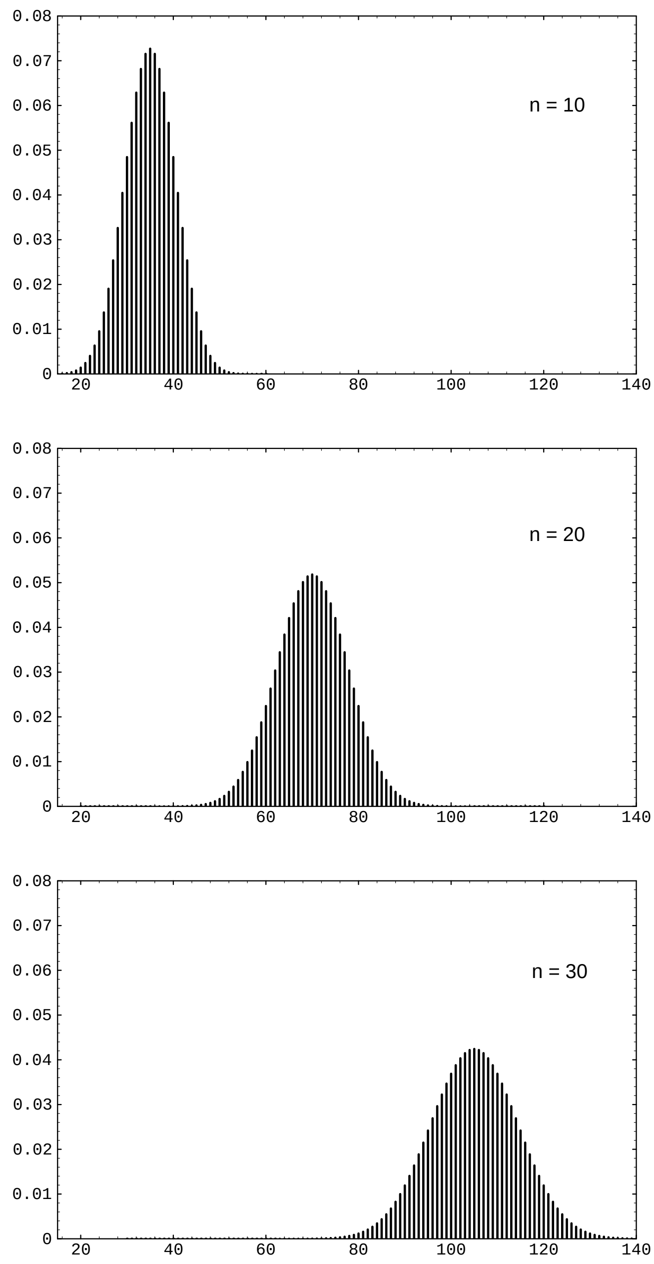

Figure 7.1: Density of S_n for rolling a die n times.

For certain special distributions it is possible to find an expression for the distribution that results from convoluting the distribution with itself n times.

The convolution of two binomial distributions, one with parameters m and p and the other with parameters n and p, is a binomial distribution with parameters $(m+n)$ and p. This fact follows easily from a consideration of the experiment which consists of first tossing a coin m times, and then tossing it n more times.

The convolution of k geometric distributions with common parameter p is a negative binomial distribution with parameters p and k. This can be seen by considering the experiment which consists of tossing a coin until the kth head appears.

Exercises

1 A die is rolled three times. Find the probability that the sum of the outcomes is

(a) greater than 9.

(b) an odd number.

2 The price of a stock on a given trading day changes according to the distribution

$$p_X = \begin{pmatrix} -1 & 0 & 1 & 2 \\ 1/4 & 1/2 & 1/8 & 1/8 \end{pmatrix}.$$

Find the distribution for the change in stock price after two (independent) trading days.

3 Let X_1 and X_2 be independent random variables with common distribution

$$p_X = \begin{pmatrix} 0 & 1 & 2 \\ 1/8 & 3/8 & 1/2 \end{pmatrix}.$$

Find the distribution of the sum $X_1 + X_2$.

4 In one play of a certain game you win an amount X with distribution

$$p_X = \begin{pmatrix} 1 & 2 & 3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}.$$

Using the program **NFoldConvolution** find the distribution for your total winnings after ten (independent) plays. Plot this distribution.

5 Consider the following two experiments: the first has outcome X taking on the values 0, 1, and 2 with equal probabilities; the second results in an (independent) outcome Y taking on the value 3 with probability 1/4 and 4 with probability 3/4. Find the distribution of

(a) $Y + X$.

(b) $Y - X$.

6 People arrive at a queue according to the following scheme: During each minute of time either 0 or 1 person arrives. The probability that 1 person arrives is p and that no person arrives is $q = 1 - p$. Let C_r be the number of customers arriving in the first r minutes. Consider a Bernoulli trials process with a success if a person arrives in a unit time and failure if no person arrives in a unit time. Let T_r be the number of failures before the rth success.

(a) What is the distribution for T_r?

(b) What is the distribution for C_r?

(c) Find the mean and variance for the number of customers arriving in the first r minutes.

7 (a) A die is rolled three times with outcomes X_1, X_2, and X_3. Let Y_3 be the maximum of the values obtained. Show that

$$P(Y_3 \leq j) = P(X_1 \leq j)^3 \ .$$

Use this to find the distribution of Y_3. Does Y_3 have a bell-shaped distribution?

(b) Now let Y_n be the maximum value when n dice are rolled. Find the distribution of Y_n. Is this distribution bell-shaped for large values of n?

8 A baseball player is to play in the World Series. Based upon his season play, you estimate that if he comes to bat four times in a game the number of hits he will get has a distribution

$$p_X = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ .4 & .2 & .2 & .1 & .1 \end{pmatrix} .$$

Assume that the player comes to bat four times in each game of the series.

(a) Let X denote the number of hits that he gets in a series. Using the program **NFoldConvolution**, find the distribution of X for each of the possible series lengths: four-game, five-game, six-game, seven-game.

(b) Using one of the distribution found in part (a), find the probability that his batting average exceeds .400 in a four-game series. (The batting average is the number of hits divided by the number of times at bat.)

(c) Given the distribution p_X, what is his long-term batting average?

9 Prove that you cannot load two dice in such a way that the probabilities for any sum from 2 to 12 are the same. (Be sure to consider the case where one or more sides turn up with probability zero.)

10 (Lévy[2]) Assume that n is an integer, not prime. Show that you can find two distributions a and b on the nonnegative integers such that the convolution of

[2]See M. Krasner and B. Ranulae, "Sur une Proprieté des Polynomes de la Division du Circle"; and the following note by J. Hadamard, in *C. R. Acad. Sci.,* vol. 204 (1937), pp. 397–399.

a and b is the equiprobable distribution on the set 0, 1, 2, ..., $n-1$. If n is prime this is not possible, but the proof is not so easy. (Assume that neither a nor b is concentrated at 0.)

11 Assume that you are playing craps with dice that are loaded in the following way: faces two, three, four, and five all come up with the same probability $(1/6)+r$. Faces one and six come up with probability $(1/6)-2r$, with $0 < r < .02$. Write a computer program to find the probability of winning at craps with these dice, and using your program find which values of r make craps a favorable game for the player with these dice.

7.2 Sums of Continuous Random Variables

In this section we consider the continuous version of the problem posed in the previous section: How are sums of independent random variables distributed?

Convolutions

Definition 7.2 Let X and Y be two continuous random variables with density functions $f(x)$ and $g(y)$, respectively. Assume that both $f(x)$ and $g(y)$ are defined for all real numbers. Then the *convolution* $f * g$ of f and g is the function given by

$$\begin{aligned}(f*g)(z) &= \int_{-\infty}^{+\infty} f(z-y)g(y)\,dy \\ &= \int_{-\infty}^{+\infty} g(z-x)f(x)\,dx\ .\end{aligned}$$

□

This definition is analogous to the definition, given in Section 7.1, of the convolution of two distribution functions. Thus it should not be surprising that if X and Y are independent, then the density of their sum is the convolution of their densities. This fact is stated as a theorem below, and its proof is left as an exercise (see Exercise 1).

Theorem 7.1 Let X and Y be two independent random variables with density functions $f_X(x)$ and $f_Y(y)$ defined for all x. Then the sum $Z = X + Y$ is a random variable with density function $f_Z(z)$, where f_Z is the convolution of f_X and f_Y. □

To get a better understanding of this important result, we will look at some examples.

Sum of Two Independent Uniform Random Variables

Example 7.3 Suppose we choose independently two numbers at random from the interval $[0, 1]$ with uniform probability density. What is the density of their sum?

Let X and Y be random variables describing our choices and $Z = X + Y$ their sum. Then we have

$$f_X(x) = f_Y(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise;} \end{cases}$$

and the density function for the sum is given by

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)\, dy\ .$$

Since $f_Y(y) = 1$ if $0 \leq y \leq 1$ and 0 otherwise, this becomes

$$f_Z(z) = \int_0^1 f_X(z-y)\, dy\ .$$

Now the integrand is 0 unless $0 \leq z - y \leq 1$ (i.e., unless $z - 1 \leq y \leq z$) and then it is 1. So if $0 \leq z \leq 1$, we have

$$f_Z(z) = \int_0^z dy = z\ ,$$

while if $1 < z \leq 2$, we have

$$f_Z(z) = \int_{z-1}^1 dy = 2 - z\ ,$$

and if $z < 0$ or $z > 2$ we have $f_Z(z) = 0$ (see Figure 7.2). Hence,

$$f_Z(z) = \begin{cases} z, & \text{if } 0 \leq z \leq 1, \\ 2 - z, & \text{if } 1 < z \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

Note that this result agrees with that of Example 2.4. $\square$

Sum of Two Independent Exponential Random Variables

Example 7.4 Suppose we choose two numbers at random from the interval $[0, \infty)$ with an *exponential* density with parameter λ. What is the density of their sum?

Let X, Y, and $Z = X + Y$ denote the relevant random variables, and f_X, f_Y, and f_Z their densities. Then

$$f_X(x) = f_Y(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise;} \end{cases}$$

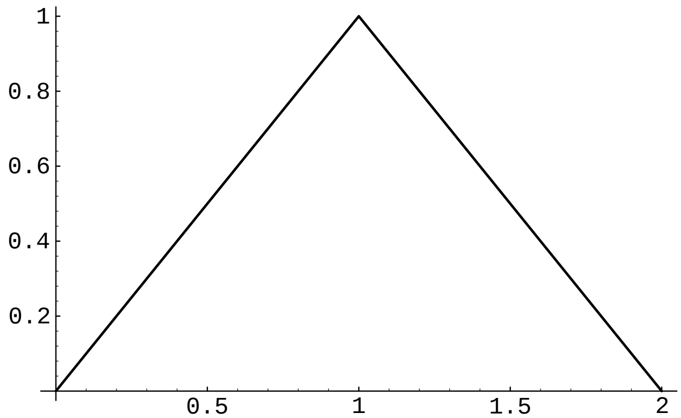

Figure 7.2: Convolution of two uniform densities.

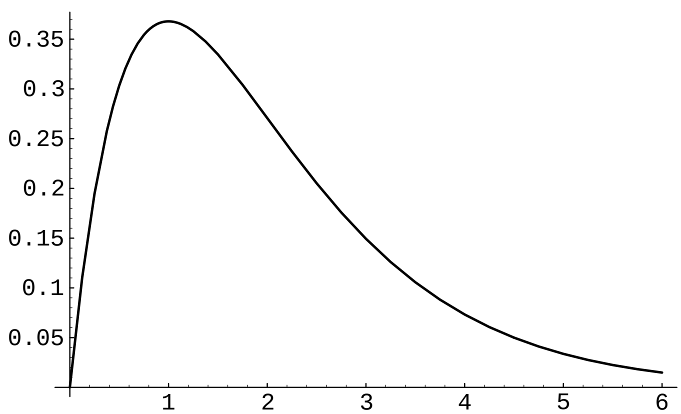

Figure 7.3: Convolution of two exponential densities with $\lambda = 1$.

and so, if $z > 0$,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)\, dy \\ &= \int_0^z \lambda e^{-\lambda(z-y)} \lambda e^{-\lambda y}\, dy \\ &= \int_0^z \lambda^2 e^{-\lambda z}\, dy \\ &= \lambda^2 z e^{-\lambda z}, \end{aligned}$$

while if $z < 0$, $f_Z(z) = 0$ (see Figure 7.3). Hence,

$$f_Z(z) = \begin{cases} \lambda^2 z e^{-\lambda z}, & \text{if } z \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

$\square$

Sum of Two Independent Normal Random Variables

Example 7.5 It is an interesting and important fact that the convolution of two normal densities with means μ_1 and μ_2 and variances σ_1 and σ_2 is again a normal density, with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. We will show this in the special case that both random variables are standard normal. The general case can be done in the same way, but the calculation is messier. Another way to show the general result is given in Example 10.17.

Suppose X and Y are two independent random variables, each with the standard *normal* density (see Example 5.8). We have

$$f_X(x) = f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} ,$$

and so

$$\begin{aligned} f_Z(z) &= f_X * f_Y(z) \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-(z-y)^2/2} e^{-y^2/2}\, dy \\ &= \frac{1}{2\pi} e^{-z^2/4} \int_{-\infty}^{+\infty} e^{-(y-z/2)^2}\, dy \\ &= \frac{1}{2\pi} e^{-z^2/4} \sqrt{\pi} \left[\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(y-z/2)^2}\, dy \right] . \end{aligned}$$

The expression in the brackets equals 1, since it is the integral of the normal density function with $\mu = 0$ and $\sigma = \sqrt{2}$. So, we have

$$f_Z(z) = \frac{1}{\sqrt{4\pi}} e^{-z^2/4} .$$

□

Sum of Two Independent Cauchy Random Variables

Example 7.6 Choose two numbers at random from the interval $(-\infty, +\infty)$ with the Cauchy density with parameter $a = 1$ (see Example 5.10). Then

$$f_X(x) = f_Y(x) = \frac{1}{\pi(1 + x^2)} ,$$

and $Z = X + Y$ has density

$$f_Z(z) = \frac{1}{\pi^2} \int_{-\infty}^{+\infty} \frac{1}{1 + (z - y)^2} \frac{1}{1 + y^2}\, dy .$$

This integral requires some effort, and we give here only the result (see Section 10.3, or Dwass[3]):

$$f_Z(z) = \frac{2}{\pi(4+z^2)} \ .$$

Now, suppose that we ask for the density function of the *average*

$$A = (1/2)(X+Y)$$

of X and Y. Then $A = (1/2)Z$. Exercise 5.2.19 shows that if U and V are two continuous random variables with density functions $f_U(x)$ and $f_V(x)$, respectively, and if $V = aU$, then

$$f_V(x) = \left(\frac{1}{a}\right) f_U\left(\frac{x}{a}\right) \ .$$

Thus, we have

$$f_A(z) = 2f_Z(2z) = \frac{1}{\pi(1+z^2)} \ .$$

Hence, the density function for the average of two random variables, each having a Cauchy density, is again a random variable with a Cauchy density; this remarkable property is a peculiarity of the Cauchy density. One consequence of this is if the error in a certain measurement process had a Cauchy density and you averaged a number of measurements, the average could not be expected to be any more accurate than any one of your individual measurements! □

Rayleigh Density

Example 7.7 Suppose X and Y are two independent standard normal random variables. Now suppose we locate a point P in the xy-plane with coordinates (X, Y) and ask: What is the density of the square of the distance of P from the origin? (We have already simulated this problem in Example 5.9.) Here, with the preceding notation, we have

$$f_X(x) = f_Y(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ .$$

Moreover, if X^2 denotes the square of X, then (see Theorem 5.1 and the discussion following)

$$\begin{aligned} f_{X^2}(r) &= \begin{cases} \frac{1}{2\sqrt{r}}(f_X(\sqrt{r}) + f_X(-\sqrt{r})) & \text{if } r > 0, \\ 0 & \text{otherwise.} \end{cases} \\ &= \begin{cases} \frac{1}{\sqrt{2\pi r}}(e^{-r/2}) & \text{if } r > 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

[3] M. Dwass, "On the Convolution of Cauchy Distributions," *American Mathematical Monthly,* vol. 92, no. 1, (1985), pp. 55–57; see also R. Nelson, letters to the Editor, ibid., p. 679.

This is a gamma density with $\lambda = 1/2$, $\beta = 1/2$ (see Example 7.4). Now let $R^2 = X^2 + Y^2$. Then

$$\begin{aligned} f_{R^2}(r) &= \int_{-\infty}^{+\infty} f_{X^2}(r-s) f_{Y^2}(s)\, ds \\ &= \frac{1}{4\pi} \int_{-\infty}^{+\infty} e^{-(r-s)/2} \frac{r-s}{2}^{-1/2} e^{-s} \frac{s}{2}^{-1/2} ds \ , \\ &= \begin{cases} \frac{1}{2} e^{-r^2/2}, & \text{if } r \ge 0, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Hence, R^2 has a gamma density with $\lambda = 1/2$, $\beta = 1$. We can interpret this result as giving the density for the square of the distance of P from the center of a target if its coordinates are normally distributed.

The density of the random variable R is obtained from that of R^2 in the usual way (see Theorem 5.1), and we find

$$f_R(r) = \begin{cases} \frac{1}{2} e^{-r^2/2} \cdot 2r = r e^{-r^2/2}, & \text{if } r \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$

Physicists will recognize this as a Rayleigh density. Our result here agrees with our simulation in Example 5.9. □

Chi-Squared Density

More generally, the same method shows that the sum of the squares of n independent normally distributed random variables with mean 0 and standard deviation 1 has a gamma density with $\lambda = 1/2$ and $\beta = n/2$. Such a density is called a *chi-squared density* with n degrees of freedom. This density was introduced in Chapter 4.3. In Example 5.10, we used this density to test the hypothesis that two traits were independent.

Another important use of the chi-squared density is in comparing experimental data with a theoretical discrete distribution, to see whether the data supports the theoretical model. More specifically, suppose that we have an experiment with a finite set of outcomes. If the set of outcomes is countable, we group them into finitely many sets of outcomes. We propose a theoretical distribution which we think will model the experiment well. We obtain some data by repeating the experiment a number of times. Now we wish to check how well the theoretical distribution fits the data.

Let X be the random variable which represents a theoretical outcome in the model of the experiment, and let $m(x)$ be the distribution function of X. In a manner similar to what was done in Example 5.10, we calculate the value of the expression

$$V = \sum_x \frac{(o_x - n \cdot m(x))^2}{n \cdot m(x)} \ ,$$

where the sum runs over all possible outcomes x, n is the number of data points, and o_x denotes the number of outcomes of type x observed in the data. Then

Outcome	Observed Frequency
1	15
2	8
3	7
4	5
5	7
6	18

Table 7.1: Observed data.

for moderate or large values of n, the quantity V is approximately chi-squared distributed, with $\nu-1$ degrees of freedom, where ν represents the number of possible outcomes. The proof of this is beyond the scope of this book, but we will illustrate the reasonableness of this statement in the next example. If the value of V is very large, when compared with the appropriate chi-squared density function, then we would tend to reject the hypothesis that the model is an appropriate one for the experiment at hand. We now give an example of this procedure.

Example 7.8 Suppose we are given a single die. We wish to test the hypothesis that the die is fair. Thus, our theoretical distribution is the uniform distribution on the integers between 1 and 6. So, if we roll the die n times, the expected number of data points of each type is $n/6$. Thus, if o_i denotes the actual number of data points of type i, for $1 \leq i \leq 6$, then the expression

$$V = \sum_{i=1}^{6} \frac{(o_i - n/6)^2}{n/6}$$

is approximately chi-squared distributed with 5 degrees of freedom.

Now suppose that we actually roll the die 60 times and obtain the data in Table 7.1. If we calculate V for this data, we obtain the value 13.6. The graph of the chi-squared density with 5 degrees of freedom is shown in Figure 7.4. One sees that values as large as 13.6 are rarely taken on by V if the die is fair, so we would reject the hypothesis that the die is fair. (When using this test, a statistician will reject the hypothesis if the data gives a value of V which is larger than 95% of the values one would expect to obtain if the hypothesis is true.)

In Figure 7.5, we show the results of rolling a die 60 times, then calculating V, and then repeating this experiment 1000 times. The program that performs these calculations is called **DieTest**. We have superimposed the chi-squared density with 5 degrees of freedom; one can see that the data values fit the curve fairly well, which supports the statement that the chi-squared density is the correct one to use. $\square$

So far we have looked at several important special cases for which the convolution integral can be evaluated explicitly. In general, the convolution of two continuous densities cannot be evaluated explicitly, and we must resort to numerical methods. Fortunately, these prove to be remarkably effective, at least for bounded densities.

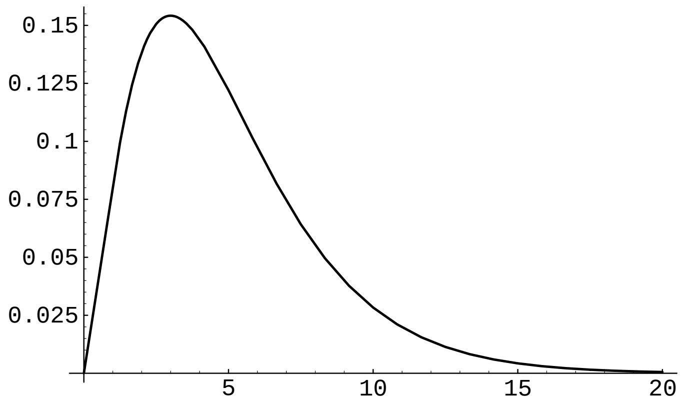

Figure 7.4: Chi-squared density with 5 degrees of freedom.

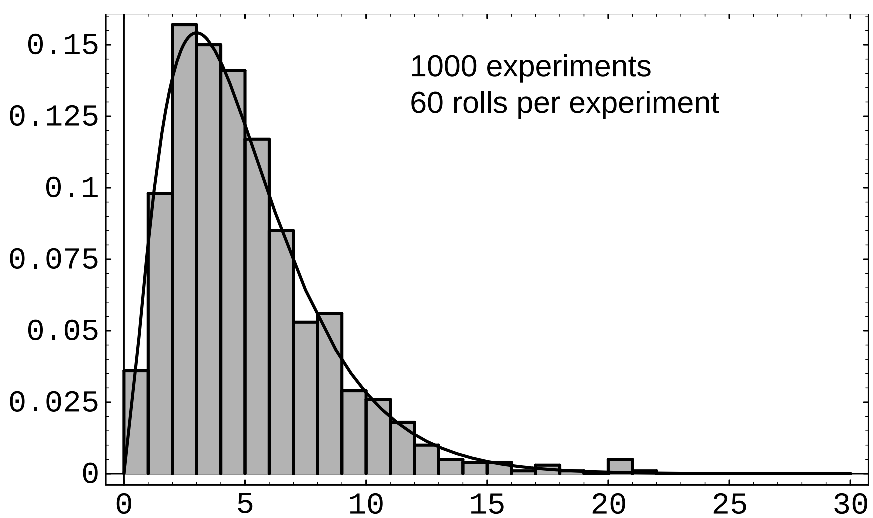

Figure 7.5: Rolling a fair die.

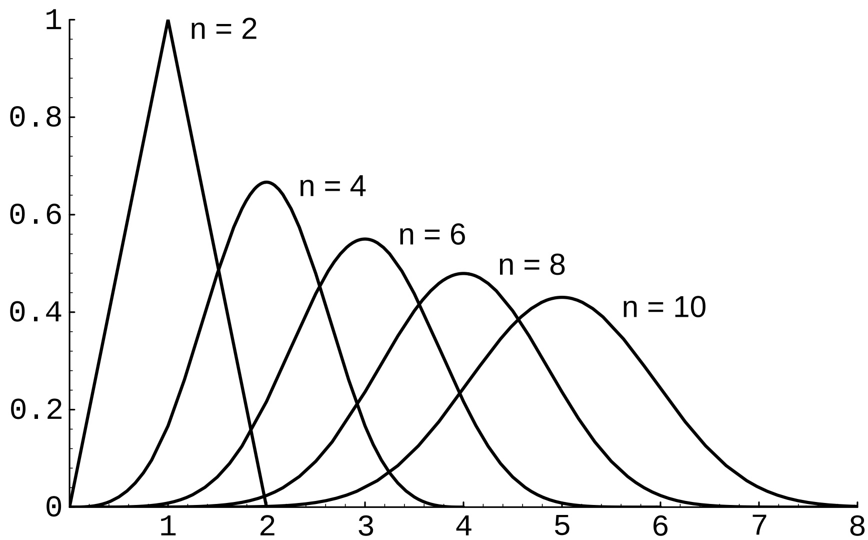

Figure 7.6: Convolution of n uniform densities.

Independent Trials

We now consider briefly the distribution of the sum of n independent random variables, all having the same density function. If X_1, X_2, $\ldots$, X_n are these random variables and $S_n = X_1 + X_2 + \cdots + X_n$ is their sum, then we will have

$$f_{S_n}(x) = (f_{X_1} * f_{X_2} * \cdots * f_{X_n})(x) ,$$

where the right-hand side is an n-fold convolution. It is possible to calculate this density for general values of n in certain simple cases.

Example 7.9 Suppose the X_i are uniformly distributed on the interval $[0, 1]$. Then

$$f_{X_i}(x) = \begin{cases} 1, & \text{if } 0 \le x \le 1, \\ 0, & \text{otherwise,} \end{cases}$$

and $f_{S_n}(x)$ is given by the formula[4]

$$f_{S_n}(x) = \begin{cases} \frac{1}{(n-1)!} \sum_{0 \le j \le x} (-1)^j \binom{n}{j} (x-j)^{n-1}, & \text{if } 0 < x < n, \\ 0, & \text{otherwise.} \end{cases}$$

The density $f_{S_n}(x)$ for $n = 2$, 4, 6, 8, 10 is shown in Figure 7.6.

If the X_i are distributed normally, with mean 0 and variance 1, then (cf. Example 7.5)

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} ,$$

[4]J. B. Uspensky, *Introduction to Mathematical Probability* (New York: McGraw-Hill, 1937), p. 277.

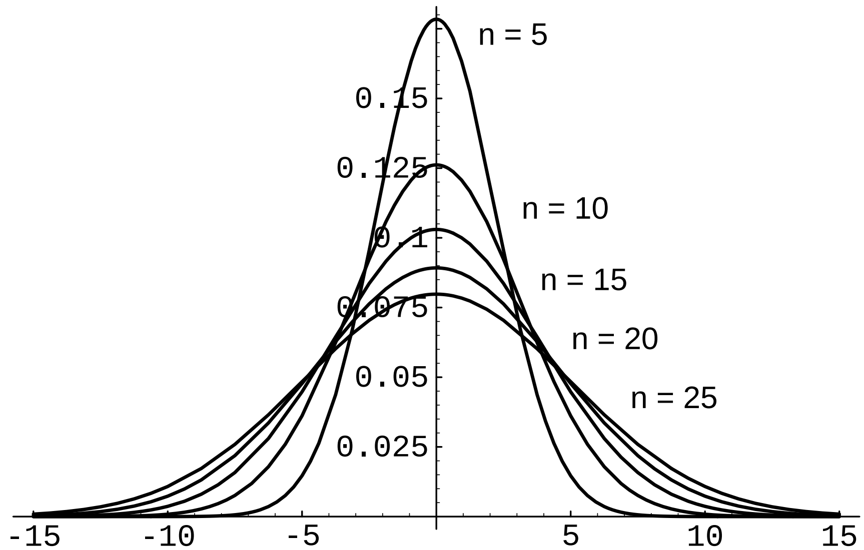

Figure 7.7: Convolution of n standard normal densities.

and

$$f_{S_n}(x) = \frac{1}{\sqrt{2\pi n}} e^{-x^2/2n} .$$

Here the density f_{S_n} for $n = 5, 10, 15, 20, 25$ is shown in Figure 7.7.

If the X_i are all exponentially distributed, with mean $1/\lambda$, then

$$f_{X_i}(x) = \lambda e^{-\lambda x} ,$$

and

$$f_{S_n}(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{n-1}}{(n-1)!} .$$

In this case the density f_{S_n} for $n = 2, 4, 6, 8, 10$ is shown in Figure 7.8. $\square$

Exercises

1 Let X and Y be independent real-valued random variables with density functions $f_X(x)$ and $f_Y(y)$, respectively. Show that the density function of the sum $X + Y$ is the convolution of the functions $f_X(x)$ and $f_Y(y)$. *Hint*: Let $\bar{X}$ be the joint random variable (X, Y). Then the joint density function of $\bar{X}$ is $f_X(x) f_Y(y)$, since X and Y are independent. Now compute the probability that $X + Y \le z$, by integrating the joint density function over the appropriate region in the plane. This gives the cumulative distribution function of Z. Now differentiate this function with respect to z to obtain the density function of z.

2 Let X and Y be independent random variables defined on the space Ω, with density functions f_X and f_Y, respectively. Suppose that $Z = X + Y$. Find the density f_Z of Z if

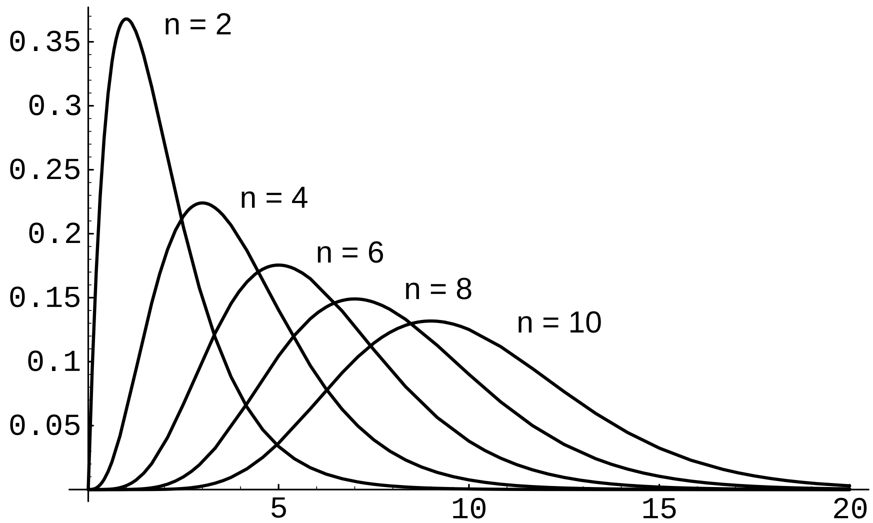

Figure 7.8: Convolution of n exponential densities with $\lambda = 1$.

(a)
$$f_X(x) = f_Y(x) = \begin{cases} 1/2, & \text{if } -1 \le x \le +1, \\ 0, & \text{otherwise.} \end{cases}$$

(b)
$$f_X(x) = f_Y(x) = \begin{cases} 1/2, & \text{if } 3 \le x \le 5, \\ 0, & \text{otherwise.} \end{cases}$$

(c)
$$f_X(x) = \begin{cases} 1/2, & \text{if } -1 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

$$f_Y(x) = \begin{cases} 1/2, & \text{if } 3 \le x \le 5, \\ 0, & \text{otherwise.} \end{cases}$$

(d) What can you say about the set $E = \{ z : f_Z(z) > 0 \}$ in each case?

3 Suppose again that $Z = X + Y$. Find f_Z if

(a)
$$f_X(x) = f_Y(x) = \begin{cases} x/2, & \text{if } 0 < x < 2, \\ 0, & \text{otherwise.} \end{cases}$$

(b)
$$f_X(x) = f_Y(x) = \begin{cases} (1/2)(x-3), & \text{if } 3 < x < 5, \\ 0, & \text{otherwise.} \end{cases}$$

(c)
$$f_X(x) = \begin{cases} 1/2, & \text{if } 0 < x < 2, \\ 0, & \text{otherwise,} \end{cases}$$

$$f_Y(x) = \begin{cases} x/2, & \text{if } 0 < x < 2, \\ 0, & \text{otherwise.} \end{cases}$$

(d) What can you say about the set $E = \{\, z : f_Z(z) > 0 \,\}$ in each case?

4 Let X, Y, and Z be independent random variables with

$$f_X(x) = f_Y(x) = f_Z(x) = \begin{cases} 1, & \text{if } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Suppose that $W = X + Y + Z$. Find f_W directly, and compare your answer with that given by the formula in Example 7.9. *Hint*: See Example 7.3.

5 Suppose that X and Y are independent and $Z = X + Y$. Find f_Z if

(a)

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

$$f_Y(x) = \begin{cases} \mu e^{-\mu x}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(b)

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

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

6 Suppose again that $Z = X + Y$. Find f_Z if

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x-\mu_1)^2/2\sigma_1^2} \\ f_Y(x) &= \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(x-\mu_2)^2/2\sigma_2^2} . \end{aligned}$$

***7** Suppose that $R^2 = X^2 + Y^2$. Find f_{R^2} and f_R if

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x-\mu_1)^2/2\sigma_1^2} \\ f_Y(x) &= \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(x-\mu_2)^2/2\sigma_2^2} . \end{aligned}$$

8 Suppose that $R^2 = X^2 + Y^2$. Find f_{R^2} and f_R if

$$f_X(x) = f_Y(x) = \begin{cases} 1/2, & \text{if } -1 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

9 Assume that the service time for a customer at a bank is exponentially distributed with mean service time 2 minutes. Let X be the total service time for 10 customers. Estimate the probability that $X > 22$ minutes.

10 Let $X_1, X_2, \ldots, X_n$ be n independent random variables each of which has an exponential density with mean μ. Let M be the *minimum* value of the X_j. Show that the density for M is exponential with mean μ/n. *Hint*: Use cumulative distribution functions.

11 A company buys 100 lightbulbs, each of which has an exponential lifetime of 1000 hours. What is the expected time for the first of these bulbs to burn out? (See Exercise 10.)

12 An insurance company assumes that the time between claims from each of its homeowners' policies is exponentially distributed with mean μ. It would like to estimate μ by averaging the times for a number of policies, but this is not very practical since the time between claims is about 30 years. At Galambos'[5] suggestion the company puts its customers in groups of 50 and observes the time of the first claim within each group. Show that this provides a practical way to estimate the value of μ.

13 Particles are subject to collisions that cause them to split into two parts with each part a fraction of the parent. Suppose that this fraction is uniformly distributed between 0 and 1. Following a single particle through several splittings we obtain a fraction of the original particle $Z_n = X_1 \cdot X_2 \cdot \ldots \cdot X_n$ where each X_j is uniformly distributed between 0 and 1. Show that the density for the random variable Z_n is

$$f_n(z) = \frac{1}{(n-1)!}(-\log z)^{n-1}.$$

Hint: Show that $Y_k = -\log X_k$ is exponentially distributed. Use this to find the density function for $S_n = Y_1 + Y_2 + \cdots + Y_n$, and from this the cumulative distribution and density of $Z_n = e^{-S_n}$.

14 Assume that X_1 and X_2 are independent random variables, each having an exponential density with parameter λ. Show that $Z = X_1 - X_2$ has density

$$f_Z(z) = (1/2)\lambda e^{-\lambda|z|} .$$

15 Suppose we want to test a coin for fairness. We flip the coin n times and record the number of times X_0 that the coin turns up tails and the number of times $X_1 = n - X_0$ that the coin turns up heads. Now we set

$$Z = \sum_{i=0}^{1} \frac{(X_i - n/2)^2}{n/2} .$$

Then for a fair coin Z has approximately a chi-squared distribution with $2 - 1 = 1$ degree of freedom. Verify this by computer simulation first for a fair coin ($p = 1/2$) and then for a biased coin ($p = 1/3$).

[5] J. Galambos, *Introductory Probability Theory* (New York: Marcel Dekker, 1984), p. 159.

16 Verify your answers in Exercise 2(a) by computer simulation: Choose X and Y from $[-1, 1]$ with uniform density and calculate $Z = X + Y$. Repeat this experiment 500 times, recording the outcomes in a bar graph on $[-2, 2]$ with 40 bars. Does the density f_Z calculated in Exercise 2(a) describe the shape of your bar graph? Try this for Exercises 2(b) and Exercise 2(c), too.

17 Verify your answers to Exercise 3 by computer simulation.

18 Verify your answer to Exercise 4 by computer simulation.

19 The *support* of a function $f(x)$ is defined to be the set

$$\{x \ : \ f(x) > 0\} \ .$$

Suppose that X and Y are two continuous random variables with density functions $f_X(x)$ and $f_Y(y)$, respectively, and suppose that the supports of these density functions are the intervals $[a, b]$ and $[c, d]$, respectively. Find the support of the density function of the random variable $X + Y$.

20 Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables, all having a common density function f_X with support $[a, b]$ (see Exercise 19). Let $S_n = X_1 + X_2 + \cdots + X_n$, with density function f_{S_n}. Show that the support of f_{S_n} is the interval $[na, nb]$. *Hint*: Write $f_{S_n} = f_{S_{n-1}} * f_X$. Now use Exercise 19 to establish the desired result by induction.

21 Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables, all having a common density function f_X. Let $A = S_n/n$ be their average. Find f_A if

(a) $f_X(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ (normal density).

(b) $f_X(x) = e^{-x}$ (exponential density).

Hint: Write $f_A(x)$ in terms of $f_{S_n}(x)$.