

## SECTION 6.4 Expected Value and Variance

*This section concludes the discussion of probability theory. The linearity of expectation is a surprisingly powerful tool, as Exercises 21 and 37, for example, illustrate.*

1. By Theorem 2 with  $p = 1/2$  and  $n = 5$ , we see that the expected number of heads is 2.5.
3. By Theorem 2 the expected number of successes for  $n$  Bernoulli trials is  $np$ . In the present problem we have  $n = 10$  and  $p = 1/6$ . Therefore the expected number of successes (i.e., appearances of a 6) is  $10 \cdot (1/6) = 1\frac{2}{3}$ .
5. This problem involves a lot of computation. It is similar to Example 3, which relied only the results of Example 12 in Section 6.2. We need to compute the probability of each outcome in order to be able to apply Definition 1 (expected value). It is easy to see that the given information implies that for one roll of such a die  $p(3) = 2/7$  and  $p(1) = p(2) = p(4) = p(5) = p(6) = 1/7$  (this was Exercise 2 in Section 6.2). Next we need to do several computations similar to those required in Exercise 5 in Section 6.2, in order to compute  $p(X = k)$  for each  $k$  from 2 to 12, where the random variable  $X$  represents the sum (for example,  $X(3, 5) = 8$ ). The probability of a sum of 2,  $p(X = 2)$ , is  $\frac{1}{7} \cdot \frac{1}{7} = \frac{1}{49}$ , since the only way to achieve a sum of 2 is to roll a 1 on each die, and the two dice are independent. Similarly,  $p(X = 3) = \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} = \frac{2}{49}$ , since both of the outcomes  $(1, 2)$  and  $(2, 1)$  give a sum of 3. We perform similar calculations for the other outcomes of the

sum. Here is the entire set of values:

$$\begin{aligned}
 p(X=2) &= \frac{1}{7} \cdot \frac{1}{7} = \frac{1}{49} \\
 p(X=3) &= \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} = \frac{2}{49} \\
 p(X=4) &= \frac{1}{7} \cdot \frac{2}{7} + \frac{1}{7} \cdot \frac{1}{7} + \frac{2}{7} \cdot \frac{1}{7} = \frac{5}{49} \\
 p(X=5) &= \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{2}{7} + \frac{2}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} = \frac{6}{49} \\
 p(X=6) &= \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} + \frac{2}{7} \cdot \frac{2}{7} + \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} = \frac{8}{49} \\
 p(X=7) &= \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} + \frac{2}{7} \cdot \frac{1}{7} + \frac{2}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} = \frac{8}{49} \\
 p(X=8) &= \frac{1}{7} \cdot \frac{1}{7} + \frac{2}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{2}{7} + \frac{1}{7} \cdot \frac{1}{7} = \frac{7}{49} \\
 p(X=9) &= \frac{2}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{2}{7} = \frac{6}{49} \\
 p(X=10) &= \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} = \frac{3}{49} \\
 p(X=11) &= \frac{1}{7} \cdot \frac{1}{7} + \frac{1}{7} \cdot \frac{1}{7} = \frac{2}{49} \\
 p(X=12) &= \frac{1}{7} \cdot \frac{1}{7} = \frac{1}{49}
 \end{aligned}$$

A check on our calculation is that the sum of the probabilities is 1. Finally, we need to add the values of  $X$  times the corresponding probabilities:

$$E(X) = 2 \cdot \frac{1}{49} + 3 \cdot \frac{2}{49} + 4 \cdot \frac{5}{49} + 5 \cdot \frac{6}{49} + 6 \cdot \frac{8}{49} + 7 \cdot \frac{8}{49} + 8 \cdot \frac{7}{49} + 9 \cdot \frac{6}{49} + 10 \cdot \frac{3}{49} + 11 \cdot \frac{2}{49} + 12 \cdot \frac{1}{49} = \frac{336}{49} \approx 6.86$$

This is a reasonable answer (you should always ask yourself if the answer is reasonable!), since the dice are not very different from ordinary dice, and with ordinary dice the expectation is 7.

7. By Theorem 3 we know that the expectation of a sum is the sum of the expectations. In the current exercise we can let  $X$  be the random variable giving the score on the true-false questions and let  $Y$  be the random variable giving the score on the multiple choice questions. In order to compute the expectation of  $X$  and of  $Y$ , let us for a moment ignore the point values, and instead just look at the number of true-false or multiple choice questions that Linda gets right. The expected number of true-false questions she gets right is the expectation of the number of successes when 50 Bernoulli trials are performed with  $p = 0.9$ . By Theorem 2 the expectation for the number of successes is  $np = 50 \cdot 0.9 = 45$ . Since each problem counts 2 points, the expectation of  $X$  is  $45 \cdot 2 = 90$ . Similarly, the expected number of multiple choice questions she gets right is the expectation of the number of successes when 25 Bernoulli trials are performed with  $p = 0.8$ , namely  $25 \cdot 0.8 = 20$ . Since each problem counts 4 points, the expectation of  $Y$  is  $20 \cdot 4 = 80$ . Therefore her expected score on the exam is  $E(X + Y) = E(X) + E(Y) = 90 + 80 = 170$ .
9. In Example 8 we found that the answer to this question when the probability that  $x$  is in the list is  $p$  is  $p(n+2) + (2n+2)(1-p)$ . Plugging in  $p = 2/3$  we have

$$\frac{2}{3} \cdot (n+2) + (2n+2) \cdot \frac{1}{3} = \frac{4n+6}{3}.$$

11. There are 10 different outcomes of our experiment (it really doesn't matter whether we get a 6 on the last roll or not). Let the random variable  $X$  be the number of times we roll the die. For  $i = 1, 2, \dots, 9$ , the probability that  $X = i$  is  $(5/6)^{i-1}(1/6)$ , since to roll the die exactly  $i$  times requires that we obtain something other than a 6 exactly  $i - 1$  times followed by a 6 on the  $i^{\text{th}}$  roll. Furthermore  $p(X = 10) = (5/6)^9$ . We need to compute  $\sum_{i=1}^{10} i \cdot p(X = i)$ . A computer algebra system gives the answer as  $50700551/10077696 \approx 5.03$ . Note that this is reasonable, since if there were no cut-off then the expected number of rolls would be 6.

13. The random variable that counts the number of rolls has a geometric distribution with  $p = 1/6$ , since the probability of getting a sum of 7 when a pair of dice is rolled is  $1/6$ . According to Theorem 4 the expected value is  $1/(1/6) = 6$ .

15. For a geometric distribution  $p(X = k) = (1 - p)^{k-1}p$  for  $k = 1, 2, 3, \dots$ . Therefore

$$p(X \geq j) = \sum_{k=j}^{\infty} p(X = k) = \sum_{k=j}^{\infty} (1 - p)^{k-1}p = p(1 - p)^{j-1} \sum_{k=0}^{\infty} (1 - p)^k = p(1 - p)^{j-1} \frac{1}{1 - (1 - p)} = (1 - p)^{j-1},$$

where we have used the formula for the sum of a geometric series at the end.

17. The random variable that counts the number of integers we need to select has a geometric distribution with  $p = 1/2302$ . According to Theorem 4 the expected value is  $1/(1/2302) = 2302$ .

19. We know from Examples 1 and 4 that  $E(X) = 7/2$  and  $E(Y) = 7$ . To compute  $E(XY)$  we need to find the value of  $XY$  averaged over the 36 equally likely possible outcomes. The following table shows the value of  $XY$ . For example, when the outcome is  $(3, 4)$ , then  $X = 3$  and  $Y = 7$ , so  $XY = 21$ .

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	6	8	10	12	14	16
3	12	15	18	21	24	27
4	20	24	28	32	36	40
5	30	35	40	45	50	55
6	42	48	54	60	66	72

We compute the average to be  $E(XY) = 329/12$ ; and  $(7/2) \cdot 7 \neq 329/12$ .

21. We follow the hint. Let  $I_j = 1$  if a run begins at the  $j^{\text{th}}$  Bernoulli trial and  $I_j = 0$  otherwise. Note that for  $I_j$  to equal 1, we must have the  $j^{\text{th}}$  trial result in  $S$  and either the  $(j - 1)^{\text{st}}$  trial result in  $F$  or  $j = 1$ . Clearly the number of runs is the sum  $R = \sum_{j=1}^n I_j$ , since exactly one  $I_j$  is 1 for each run. Now  $E(I_1) = p$ , the probability of  $S$  on the first trial, and  $E(I_j) = p(1 - p)$  for  $1 < j \leq n$  since we need success on the  $j^{\text{th}}$  trial and failure on the  $(j - 1)^{\text{st}}$  trial. By linearity (Theorem 3, which applies even when these random variables are not independent, which they certainly are not here) we have  $E(R) = p + \sum_{j=2}^n p(1 - p) = p + (n - 1)p(1 - p)$ .

23. In Example 18 we saw that the variance of the number of successes in  $n$  Bernoulli trials is  $npq$ . Here  $n = 10$  and  $p = q = 1/2$ . Therefore the variance is  $5/2$ . Note that the unit for stating variance is not flips here, but  $(\text{flips})^2$ . To restate this in terms of flips, we must take the square root to compute the standard deviation. The standard deviation is  $\sqrt{2.5} \approx 1.6$  flips.

25. The question is asking about the signed difference. For example, if  $n = 6$  and we get five tails and one head, then  $X_6 = 4$ , whereas if we get five heads and one tail, then  $X_6 = -4$ . The key here is to notice that  $X_n$  is just  $n$  minus twice the number of heads.

a) The expected number of heads is  $n/2$ . Therefore the expected value of twice the number of heads is twice this, or  $n$ , and the expected value of  $n$  minus this is  $n - n = 0$ . This is not surprising; if it were not zero, then there would be a bias favoring heads or favoring tails.

b) Since the expected value is 0, the variance is the expected value of the square of  $X_n$ , which is the same as the square of twice the number of heads. This is clearly four times the square of the number of heads, so its expected value is  $4 \cdot n/4 = n$ , from Example 18, since  $p = q = 1/2$ . Therefore the answer is  $n$ . We are implicitly using here the result of Exercise 25 in the supplementary exercises for this chapter.

27. a) The probabilities that  $(X_1, X_2, X_3) = (x_1, x_2, x_3)$  are as follows, because with probability 1 we have that  $X_3 = (X_1 + X_2) \bmod 2$ :

$$\begin{aligned} p(0, 0, 0) &= \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = \frac{1}{4} \\ p(0, 0, 1) &= \frac{1}{2} \cdot \frac{1}{2} \cdot 0 = 0 \\ p(0, 1, 0) &= \frac{1}{2} \cdot \frac{1}{2} \cdot 0 = 0 \\ p(0, 1, 1) &= \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = \frac{1}{4} \\ p(1, 0, 0) &= \frac{1}{2} \cdot \frac{1}{2} \cdot 0 = 0 \\ p(1, 0, 1) &= \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = \frac{1}{4} \\ p(1, 1, 0) &= \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = \frac{1}{4} \\ p(1, 1, 1) &= \frac{1}{2} \cdot \frac{1}{2} \cdot 0 = 0 \end{aligned}$$

We must show that  $X_1$  and  $X_2$  are independent, that  $X_1$  and  $X_3$  are independent, and that  $X_2$  and  $X_3$  are independent. We are told that  $X_1$  and  $X_2$  are independent. To see that  $X_1$  and  $X_3$  are independent, we note from the list above that  $p(X_1 = 0 \wedge X_3 = 0) = 1/4 + 0 = 1/4$ , that  $p(X_1 = 0) = 1/2$ , and that  $p(X_3 = 0) = 1/2$ , so it is true that  $p(X_1 = 0 \wedge X_3 = 0) = p(X_1 = 0)p(X_3 = 0)$ . Essentially the same calculation shows that  $p(X_1 = 0 \wedge X_3 = 1) = p(X_1 = 0)p(X_3 = 1)$ ,  $p(X_1 = 1 \wedge X_3 = 0) = p(X_1 = 1)p(X_3 = 0)$ , and  $p(X_1 = 1 \wedge X_3 = 1) = p(X_1 = 1)p(X_3 = 1)$ . Therefore by definition,  $X_1$  and  $X_3$  are independent. The same reasoning shows that  $X_2$  and  $X_3$  are independent. To see that  $X_3$  and  $X_1 + X_2$  are not independent, we observe that  $p(X_3 = 1 \wedge X_1 + X_2 = 2) = 0$ , because  $2 \bmod 2 \neq 1$ . But  $p(X_3 = 1)p(X_1 + X_2 = 2) = (1/2)(1/4) = 1/8$ .

b) We see from the table in part (a) that  $X_1$ ,  $X_2$ , and  $X_3$  are all Bernoulli random variables, so the variance of each is  $(1/2)(1/2) = 1/4$  by Example 14. Therefore  $V(X_1) + V(X_2) + V(X_3) = 3/4$ . We see from the table above that  $p(X_1 + X_2 + X_3 = 0) = 1/4$ ,  $p(X_1 + X_2 + X_3 = 1) = 0$ ,  $p(X_1 + X_2 + X_3 = 2) = 3/4$ , and  $p(X_1 + X_2 + X_3 = 3) = 0$ . Therefore the expected value of  $X_1 + X_2 + X_3$  is  $(1/4)(0) + (3/4)(2) = 3/2$ , and the variance of  $X_1 + X_2 + X_3$  is  $(1/4)(0 - 3/2)^2 + (3/4)(2 - 3/2)^2 = 3/4$ .

c) If we attempt to prove this by mathematical induction, then presumably we would like the inductive step to be  $V((X_1 + X_2 + \cdots + X_k) + X_{k+1}) = V(X_1 + X_2 + \cdots + X_k) + V(X_{k+1})$  (by the  $n = 2$  case of Theorem 7, which was proved in the text), which then equals  $(V(X_1) + V(X_2) + \cdots + V(X_k)) + V(X_{k+1})$  by the inductive hypothesis. However, in order to invoke Theorem 7, we must have that  $X_1 + X_2 + \cdots + X_k$  and  $X_{k+1}$  are independent, and we see from part (a) that this is not a valid conclusion if all we know is the *pairwise* independence of the variables. Notice that the conclusion (that the variance of the sum is the sum of variances) is true assuming only pairwise independence; it's just that we cannot prove it in this manner. See Exercise 28.

29. We proceed as in Example 19, applying Chebyshev's inequality with  $V(X) = n/4$  by Example 18 and  $r = 5\sqrt{n}$ . We have  $p(|X(s) - E(X)| \geq 5\sqrt{n}) \leq V(X)/r^2 = (n/4)/(5\sqrt{n})^2 = 1/100$ .
31. For simplicity we suppress the argument and write simply  $X$  for  $X(s)$ . As in the proof of Theorem 5,  $E(X) = \sum_r r \cdot p(X = r)$ . Dividing both sides by  $a$  we obtain  $E(X)/a = \sum_r (r/a) \cdot p(X = r)$ . Now this sum

is at least as great as the subsum restricted to those values of  $r \geq a$ , and for those values,  $r/a \geq 1$ . Thus we have  $E(X)/a \geq \sum_{r \geq a} 1 \cdot p(X = r)$ . But this latter expression is just  $p(X \geq a)$ , as desired.

33. It is interesting to note that Markov was Chebyshev's student in Russia. One note of caution—the variance is not 2,500 cans; it is 2,500 square cans (the units for the variance of  $X$  are the square of the units for  $X$ ). So a measure of how much the number of cans recycled per day varies is about the square root of this, or about 50 cans.

a) We have  $E(X) = 50,000$  and we take  $a = 55,000$ . Then  $p(X \geq 55,000) \leq 50,000/55,000 = 10/11$ . This is not a terribly good estimate.

b) We apply Theorem 8, with  $r = 10,000$ . The probability that the number of cans recycled will differ from the expectation of 50,000 by at least 10,000 is at most  $10,000/2,500^2 = 0.0016$ . Therefore the probability is at least 0.9984 that the center will recycle between 40,000 and 60,000 cans. This is also not a very good estimate, since if the number of cans recycled per day usually differs by only about 50 from the mean of 50,000, it is virtually impossible that the difference would ever be over 200 times this amount—the probability is much, much less than 1.6 in 1000.

35. a) Each of the  $n!$  permutations occurs with probability  $1/n!$ , so clearly  $E(X)$  is the average number of comparisons, averaged over all these permutations.

b) In Example 5 of Section 3.3, we noted that the version of bubble sort that continues  $n-1$  rounds regardless of whether new changes were made uses  $n(n-1)/2$  comparisons, so  $X$  in this problem is always at most  $n(n-1)/2$ . It follows from the formula for expectation that  $E(X) \leq n(n-1)/2$ .

c) An inversion in a permutation is a pair of integers  $a_j$  and  $a_k$  with  $j < k$  (so that  $a_j < a_k$ ) such that  $a_k$  precedes  $a_j$  in the permutation (the elements are out of order). Because the bubble sort works by comparing adjacent elements and then swapping them if they are out of order, the only way that these elements can end up in their correct positions is if they are swapped, and therefore they must be compared.

d) For each permutation  $P$ , we know from part (c) that  $X(P) \geq I(P)$ . It follows from the definition of expectation that  $E(X) \geq E(I)$ .

e) This summation just counts 1 for every instance of an inversion.

f) This follows from the linearity of expectation (Theorem 3).

g) By Theorem 2 with  $n = 1$ , the expectation of  $I_{j,k}$  is the probability that  $a_k$  precedes  $a_j$  in the permutation. But by symmetry, since the permutation is randomly chosen, this is clearly  $1/2$ . (There are  $n!/2$  permutations in which  $a_k$  precedes  $a_j$  and  $n!/2$  permutations in which  $a_j$  precedes  $a_k$  out of the  $n!$  permutations in all.)

h) The summation in part (f) consists of  $C(n, 2) = n(n-1)/2$  terms, each equal to  $1/2$ , so the sum is  $(n(n-1)/2)(1/2) = n(n-1)/4$ .

i) From part (b) we know that  $E(X)$ , the object of interest, is at most  $n(n-1)/2$ , and from part (d) and part (h) we know that  $E(X)$  is at least  $n(n-1)/4$ . Since both of these are  $\Theta(n^2)$ , the result follows.

37. Following the hint, we let  $X = X_1 + X_2 + \cdots + X_n$ , where  $X_i = 1$  if the permutation fixes the  $i^{\text{th}}$  element and  $X_i = 0$  otherwise. Then  $X$  is the number of fixed elements, so we are being asked to compute  $V(X)$ . By Theorem 6, it suffices to compute  $E(X^2) - E(X)^2$ . Now  $E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = n \cdot (1/n) = 1$ , since the probability that the  $i^{\text{th}}$  element stays in its original position is clearly  $1/n$  by symmetry. To compute  $E(X^2)$  we first multiply out:

$$X^2 = (X_1 + X_2 + \cdots + X_n)^2 = \sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j$$

Now since  $X_i^2 = X_i$ ,  $E(X_i^2) = 1/n$ . Next note that  $X_i X_j = 1$  if both the  $i^{\text{th}}$  and the  $j^{\text{th}}$  elements are left fixed, and since there are  $(n-2)!$  permutations that do this, the probability that  $X_i X_j = 1$  is

$(n-2)!/n! = 1/(n(n-1))$ . Of course  $X_i X_j = 0$  otherwise. Therefore  $E(X_i X_j) = 1/(n(n-1))$ . Note also that there are  $n$  summands in the first sum of the displayed equation and  $n(n-1)$  in the second. So  $E(X^2) = n(1/n) + n(n-1) \cdot 1/(n(n-1)) = 1 + 1 = 2$ . Therefore  $V(X) = E(X^2) - E(X)^2 = 2 - 1^2 = 1$ .

39. We can prove this by doing some algebra, using Theorems 3 and 6 and Exercise 38:

$$\begin{aligned} V(X+Y) &= E((X+Y)^2) - E(X+Y)^2 \\ &= E(X^2 + 2X \cdot Y + Y^2) - (E(X) + E(Y))^2 \\ &= E(X^2) + 2E(X \cdot Y) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\ &= (E(X^2) - E(X)^2) + 2(E(X \cdot Y) - E(X)E(Y)) + (E(Y^2) - E(Y)^2) \\ &= V(X) + 2\text{Cov}(X, Y) + V(Y) \end{aligned}$$

41. The probability that a particular ball fails to go into the first bin is  $(n-1)/n$ . Since these choices are assumed to be made independently, the probability that the first bin remains empty is then  $((n-1)/n)^m$ .

43. Let  $X = X_1 + X_2 + \cdots + X_n$ , where  $X_i = 1$  if the  $i^{\text{th}}$  bin remains empty and  $X_i = 0$  otherwise. Then  $X$  is the number of bins that remain empty, so we are being asked to compute  $E(X)$ . From Exercise 41 we know that  $p(X_i = 1) = ((n-1)/n)^m$  if  $m$  balls are distributed, so  $E(X_i) = ((n-1)/n)^m$ . By linearity of expectation (Theorem 3), the expected number of bins that remain empty is therefore  $n((n-1)/n)^m = (n-1)^m/n^{m-1}$ .