Improving Your Lottery Chances

State lotteries have become very popular in America. People spend millions of dollars each week to purchase tickets with very small chances of winning medium to enormous prizes. With so much money being spent on lottery tickets, it should not be surprising that a few enterprising individuals have concocted schemes to cash in on the probabilistic naïveté of the ticket-buying public. There are now several books and videos available that claim to help lottery players improve their performance. People actually pay money for these items. Some of the advice is just common sense, but some of it is misleading and plays on subtle misconceptions about probability.

For concreteness, suppose that we have a game in which there are 40 balls numbered 1 to 40 and six are drawn without replacement to determine the winning combination. A ticket purchase requires the customer to choose six different numbers from 1 to 40 and pay a fee. This game has $\binom{40}{6} = 3{,}838{,}380$ different winning combinations and the same number of possible tickets. One piece of advice often found in published lottery aids is not to choose the six numbers on your ticket too far apart. Many people tend to pick their six numbers uniformly spread out from 1 to 40, but the winning combination often has two consecutive numbers or at least two numbers very close together. Some of these "advisors" recommend that, since it is more likely that there will be numbers close together, players should bunch some of their six numbers close together. Such advice might make sense in order to avoid choosing the same numbers as other players in a parimutuel game (i.e., a game in which all winners share the jackpot). But the idea that any strategy can improve your chances of winning is misleading.

To see why this advice is misleading, let E be the event that the winning combination contains at least one pair of consecutive numbers. The reader can calculate $\Pr(E)$ in Exercise 13 in Sec. 1.12. For this example, $\Pr(E) = 0.577$. So the lottery aids are correct that E has high probability. However, by claiming that choosing a ticket in E increases your chance of winning, they confuse the probability of the event E with the probability of each outcome in E. If you choose the ticket (5, 7, 14, 23, 24, 38), your probability of winning is only 1/3,828,380, just as it would be if you chose any other ticket. The fact that this ticket happens to be in E doesn't make your probability of winning equal to 0.577. The reason that $\Pr(E)$ is so big is that so many different combinations are in E. Each of those combinations still has probability 1/3,828,380 of winning, and you only get one combination on each ticket. The fact that there are so many combinations in E does not make each one any more likely than anything else.

1.12 Supplementary Exercises

1. Suppose that a coin is tossed seven times. Let A denote the event that a head is obtained on the first toss, and let B denote the event that a head is obtained on the fifth toss. Are A and B disjoint?

2. If A, B, and D are three events such that $\Pr(A \cup B \cup D) = 0.7$, what is the value of $\Pr(A^c \cap B^c \cap D^c)$?

3. Suppose that a certain precinct contains 350 voters, of which 250 are Democrats and 100 are Republicans. If 30 voters are chosen at random from the precinct, what is the probability that exactly 18 Democrats will be selected?

4. Suppose that in a deck of 20 cards, each card has one of the numbers 1, 2, 3, 4, or 5 and there are four cards with each number. If 10 cards are chosen from the deck at random, without replacement, what is the probability that each of the numbers 1, 2, 3, 4, and 5 will appear exactly twice?

5. Consider the contractor in Example 1.5.4 on page 19. He wishes to compute the probability that the total utility demand is high, meaning that the sum of water and electrical demand (in the units of Example 1.4.5) is at least

215. Draw a picture of this event on a graph like Fig. 1.5 or Fig. 1.9 and find its probability.

6. Suppose that a box contains r red balls and w white balls. Suppose also that balls are drawn from the box one at a time, at random, without replacement. **(a)** What is the probability that all r red balls will be obtained before any white balls are obtained? **(b)** What is the probability that all r red balls will be obtained before two white balls are obtained?

7. Suppose that a box contains r red balls, w white balls, and b blue balls. Suppose also that balls are drawn from the box one at a time, at random, without replacement. What is the probability that all r red balls will be obtained before any white balls are obtained?

8. Suppose that 10 cards, of which seven are red and three are green, are put at random into 10 envelopes, of which seven are red and three are green, so that each envelope contains one card. Determine the probability that exactly k envelopes will contain a card with a matching color ($k = 0, 1, \ldots, 10$).

9. Suppose that 10 cards, of which five are red and five are green, are put at random into 10 envelopes, of which seven are red and three are green, so that each envelope contains one card. Determine the probability that exactly k envelopes will contain a card with a matching color ($k = 0, 1, \ldots, 10$).

10. Suppose that the events A and B are disjoint. Under what conditions are A^c and B^c disjoint?

11. Let A_1, A_2, and A_3 be three arbitrary events. Show that the probability that exactly one of these three events will occur is

$$\Pr(A_1) + \Pr(A_2) + \Pr(A_3) - 2\Pr(A_1 \cap A_2) - 2\Pr(A_1 \cap A_3) - 2\Pr(A_2 \cap A_3) + 3\Pr(A_1 \cap A_2 \cap A_3).$$

12. Let $A_1, \ldots, A_n$ be n arbitrary events. Show that the probability that exactly one of these n events will occur is

$$\sum_{i=1}^{n} \Pr(A_i) - 2\sum_{i<j} \Pr(A_i \cap A_j) + 3\sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} n \Pr(A_1 \cap A_2 \cdots \cap A_n).$$

13. Consider a state lottery game in which each winning combination and each ticket consists of one set of k numbers chosen from the numbers 1 to n without replacement. We shall compute the probability that the winning combination contains at least one pair of consecutive numbers.

a. Prove that if $n < 2k - 1$, then every winning combination has at least one pair of consecutive numbers. For the rest of the problem, assume that $n \ge 2k - 1$.

b. Let $i_1 < \cdots < i_k$ be an arbitrary possible winning combination arranged in order from smallest to largest. For $s = 1, \ldots, k$, let $j_s = i_s - (s - 1)$. That is,

$$\begin{aligned} j_1 &= i_1, \\ j_2 &= i_2 - 1 \\ &\vdots \\ j_k &= i_k - (k - 1). \end{aligned}$$

Prove that $(i_1, \ldots, i_k)$ contains at least one pair of consecutive numbers if and only if $(j_1, \ldots, j_k)$ contains repeated numbers.

c. Prove that $1 \le j_1 \le \cdots \le j_k \le n - k + 1$ and that the number of $(j_1, \ldots, j_k)$ sets with no repeats is $\binom{n-k+1}{k}$.

d. Find the probability that there is no pair of consecutive numbers in the winning combination.

e. Find the probability of at least one pair of consecutive numbers in the winning combination.

5. Suppose that on each play of a certain game, a person is equally likely to win one dollar or lose one dollar. Suppose also that the person's goal is to win two dollars by playing this game. How large an initial fortune must the person have in order for the probability to be at least 0.99 that she will achieve her goal before she loses her initial fortune?

6. Suppose that on each play of a certain game, a person will either win one dollar with probability 2/3 or lose one dollar with probability 1/3. Suppose also that the person's goal is to win two dollars by playing this game. How large an initial fortune must the person have in order for the probability to be at least 0.99 that he will achieve his goal before he loses his initial fortune?

7. Suppose that on each play of a certain game, a person will either win one dollar with probability 1/3 or lose one dollar with probability 2/3. Suppose also that the person's goal is to win two dollars by playing this game. Show that no matter how large the person's initial fortune might be, the probability that she will achieve her goal before she loses her initial fortune is less than 1/4.

8. Suppose that the probability of a head on any toss of a certain coin is p $(0 < p < 1)$, and suppose that the coin is tossed repeatedly. Let X_n denote the total number of heads that have been obtained on the first n tosses, and let $Y_n = n - X_n$ denote the total number of tails on the first n tosses. Suppose that the tosses are stopped as soon as a number n is reached such that either $X_n = Y_n + 3$ or $Y_n = X_n + 3$. Determine the probability that $X_n = Y_n + 3$ when the tosses are stopped.

9. Suppose that a certain box A contains five balls and another box B contains 10 balls. One of these two boxes is selected at random, and one ball from the selected box is transferred to the other box. If this process of selecting a box at random and transferring one ball from that box to the other box is repeated indefinitely, what is the probability that box A will become empty before box B becomes empty?

2.5 Supplementary Exercises

1. Suppose that A, B, and D are any three events such that $\Pr(A|D) \ge \Pr(B|D)$ and $\Pr(A|D^c) \ge \Pr(B|D^c)$. Prove that $\Pr(A) \ge \Pr(B)$.

2. Suppose that a fair coin is tossed repeatedly and independently until both a head and a tail have appeared at least once. **(a)** Describe the sample space of this experiment. **(b)** What is the probability that exactly three tosses will be required?

3. Suppose that A and B are events such that $\Pr(A) = 1/3$, $\Pr(B) = 1/5$, and $\Pr(A|B) + \Pr(B|A) = 2/3$. Evaluate $\Pr(A^c \cup B^c)$.

4. Suppose that A and B are independent events such that $\Pr(A) = 1/3$ and $\Pr(B) > 0$. What is the value of $\Pr(A \cup B^c|B)$?

5. Suppose that in 10 rolls of a balanced die, the number 6 appeared exactly three times. What is the probability that the first three rolls each yielded the number 6?

6. Suppose that A, B, and D are events such that A and B are independent, $\Pr(A \cap B \cap D) = 0.04$, $\Pr(D|A \cap B) = 0.25$, and $\Pr(B) = 4\Pr(A)$. Evaluate $\Pr(A \cup B)$.

7. Suppose that the events A, B, and C are mutually independent. Under what conditions are A^c, B^c, and C^c mutually independent?

8. Suppose that the events A and B are disjoint and that each has positive probability. Are A and B independent?

9. Suppose that A, B, and C are three events such that A and B are disjoint, A and C are independent, and B and C are independent. Suppose also that $4\Pr(A) = 2\Pr(B) = \Pr(C) > 0$ and $\Pr(A \cup B \cup C) = 5\Pr(A)$. Determine the value of $\Pr(A)$.

10. Suppose that each of two dice is loaded so that when either die is rolled, the probability that the number k will appear is 0.1 for $k = 1, 2, 5$, or 6 and is 0.3 for $k = 3$ or 4. If the two loaded dice are rolled independently, what is the probability that the sum of the two numbers that appear will be 7?

11. Suppose that there is a probability of 1/50 that you will win a certain game. If you play the game 50 times, independently, what is the probability that you will win at least once?

12. Suppose that a balanced die is rolled three times, and let X_i denote the number that appears on the ith roll $(i = 1, 2, 3)$. Evaluate $\Pr(X_1 > X_2 > X_3)$.

13. Three students A, B, and C are enrolled in the same class. Suppose that A attends class 30 percent of the time, B attends class 50 percent of the time, and C attends class 80 percent of the time. If these students attend class independently of each other, what is (a) the probability that at least one of them will be in class on a particular day and (b) the probability that exactly one of them will be in class on a particular day?

14. Consider the World Series of baseball, as described in Exercise 16 of Sec. 2.2. If there is probability p that team A will win any particular game, what is the probability that it will be necessary to play seven games in order to determine the winner of the Series?

15. Suppose that three red balls and three white balls are thrown at random into three boxes and and that all throws are independent. What is the probability that each box contains one red ball and one white ball?

16. If five balls are thrown at random into n boxes, and all throws are independent, what is the probability that no box contains more than two balls?

17. Bus tickets in a certain city contain four numbers, U, V, W, and X. Each of these numbers is equally likely to be any of the 10 digits $0, 1, \ldots, 9$, and the four numbers are chosen independently. A bus rider is said to be lucky if $U + V = W + X$. What proportion of the riders are lucky?

18. A certain group has eight members. In January, three members are selected at random to serve on a committee. In February, four members are selected at random and independently of the first selection to serve on another committee. In March, five members are selected at random and independently of the previous two selections to serve on a third committee. Determine the probability that each of the eight members serves on at least one of the three committees.

19. For the conditions of Exercise 18, determine the probability that two particular members A and B will serve together on at least one of the three committees.

20. Suppose that two players A and B take turns rolling a pair of balanced dice and that the winner is the first player who obtains the sum of 7 on a given roll of the two dice. If A rolls first, what is the probability that B will win?

21. Three players A, B, and C take turns tossing a fair coin. Suppose that A tosses the coin first, B tosses second, and C tosses third; and suppose that this cycle is repeated indefinitely until someone wins by being the first player to obtain a head. Determine the probability that each of three players will win.

22. Suppose that a balanced die is rolled repeatedly until the same number appears on two successive rolls, and let X denote the number of rolls that are required. Determine the value of $\Pr(X = x)$, for $x = 2, 3, \ldots$.

23. Suppose that 80 percent of all statisticians are shy, whereas only 15 percent of all economists are shy. Suppose also that 90 percent of the people at a large gathering are economists and the other 10 percent are statisticians. If you meet a shy person at random at the gathering, what is the probability that the person is a statistician?

24. Dreamboat cars are produced at three different factories A, B, and C. Factory A produces 20 percent of the total output of Dreamboats, B produces 50 percent, and C produces 30 percent. However, 5 percent of the cars produced at A are lemons, 2 percent of those produced at B are lemons, and 10 percent of those produced at C are lemons. If you buy a Dreamboat and it turns out to be a lemon, what is the probability that it was produced at factory A?

25. Suppose that 30 percent of the bottles produced in a certain plant are defective. If a bottle is defective, the probability is 0.9 that an inspector will notice it and remove it from the filling line. If a bottle is not defective, the probability is 0.2 that the inspector will think that it is defective and remove it from the filling line.

a. If a bottle is removed from the filling line, what is the probability that it is defective?

b. If a customer buys a bottle that has not been removed from the filling line, what is the probability that it is defective?

26. Suppose that a fair coin is tossed until a head is obtained and that this entire experiment is then performed independently a second time. What is the probability that the second experiment requires more tosses than the first experiment?

27. Suppose that a family has exactly n children ($n \geq 2$). Assume that the probability that any child will be a girl is 1/2 and that all births are independent. Given that the family has at least one girl, determine the probability that the family has at least one boy.

28. Suppose that a fair coin is tossed independently n times. Determine the probability of obtaining exactly $n - 1$ heads, given **(a)** that at least $n - 2$ heads are obtained and **(b)** that heads are obtained on the first $n - 2$ tosses.

29. Suppose that 13 cards are selected at random from a regular deck of 52 playing cards.

a. If it is known that at least one ace has been selected, what is the probability that at least two aces have been selected?

b. If it is known that the ace of hearts has been selected, what is the probability that at least two aces have been selected?

30. Suppose that n letters are placed at random in n envelopes, as in the matching problem of Sec. 1.10, and let q_n denote the probability that no letter is placed in the correct envelope. Show that the probability that exactly one letter is placed in the correct envelope is q_{n-1}.

31. Consider again the conditions of Exercise 30. Show that the probability that exactly two letters are placed in the correct envelopes is $(1/2)q_{n-2}$.

32. Consider again the conditions of Exercise 7 of Sec. 2.2. If exactly one of the two students A and B is in class on a given day, what is the probability that it is A?

33. Consider again the conditions of Exercise 2 of Sec. 1.10. If a family selected at random from the city chosen on her previous purchase, and the probability is 2/3 that she will switch brands.

a. If her first purchase is brand A, what is the probability that her fifth purchase will be brand B?

b. If her first purchase is brand B, what is the probability that her fifth purchase will be brand B?

12. Suppose that three boys A, B, and C are throwing a ball from one to another. Whenever A has the ball, he throws it to B with a probability of 0.2 and to C with a probability of 0.8. Whenever B has the ball, he throws it to A with a probability of 0.6 and to C with a probability of 0.4. Whenever C has the ball, he is equally likely to throw it to either A or B.

a. Consider this process to be a Markov chain and construct the transition matrix.

b. If each of the three boys is equally likely to have the ball at a certain time n, which boy is most likely to have the ball at time $n+2$?

13. Suppose that a coin is tossed repeatedly in such a way that heads and tails are equally likely to appear on any given toss and that all tosses are independent, with the following exception: Whenever either three heads or three tails have been obtained on three successive tosses, then the outcome of the next toss is always of the opposite type. At time n ($n \geq 3$), let the state of this process be specified by the outcomes on tosses $n-2$, $n-1$, and n. Show that this process is a Markov chain with stationary transition probabilities and construct the transition matrix.

14. There are two boxes A and B, each containing red and green balls. Suppose that box A contains one red ball and two green balls and box B contains eight red balls and two green balls. Consider the following process: One ball is selected at random from box A, and one ball is selected at random from box B. The ball selected from box A is then placed in box B and the ball selected from box B is placed in box A. These operations are then repeated indefinitely. Show that the numbers of red balls in box A form a Markov chain with stationary transition probabilities, and construct the transition matrix of the Markov chain.

15. Verify the rows of the transition matrix in Example 3.10.6 that correspond to current states $\{AA, Aa\}$ and $\{Aa, aa\}$.

16. Let the initial probability vector in Example 3.10.6 be $v = (1/16, 1/4, 1/8, 1/4, 1/4, 1/16)$. Find the probabilities of the six states after one generation.

17. Return to Example 3.10.6. Assume that the state at time $n-1$ is $\{Aa, aa\}$.

a. Suppose that we learn that X_{n+1} is $\{AA, aa\}$. Find the conditional distribution of X_n. (That is, find all the probabilities for the possible states at time n given that the state at time $n+1$ is $\{AA, aa\}$.)

b. Suppose that we learn that X_{n+1} is $\{aa, aa\}$. Find the conditional distribution of X_n.

18. Return to Example 3.10.13. Prove that the stationary distributions described there are the only stationary distributions for that Markov chain.

19. Find the unique stationary distribution for the Markov chain in Exercise 2.

20. The unique stationary distribution in Exercise 9 is $\mathbf{v} = (0, 1, 0, 0)$. This is an instance of the following general result: Suppose that a Markov chain has exactly one absorbing state. Suppose further that, for each non-absorbing state k, there is n such that the probability is positive of moving from state k to the absorbing state in n steps. Then the unique stationary distribution has probability 1 in the absorbing state. Prove this result.

3.11 Supplementary Exercises

1. Suppose that X and Y are independent random variables, that X has the uniform distribution on the integers 1, 2, 3, 4, 5 (discrete), and that Y has the uniform distribution on the interval [0, 5] (continuous). Let Z be a random variable such that $Z = X$ with probability 1/2 and $Z = Y$ with probability 1/2. Sketch the c.d.f. of Z.

2. Suppose that X and Y are independent random variables. Suppose that X has a discrete distribution concentrated on finitely many distinct values with p.f. f_1. Suppose that Y has a continuous distribution with p.d.f. f_2. Let $Z = X + Y$. Show that Z has a continuous distribution and find its p.d.f. *Hint:* First find the conditional p.d.f. of Z given $X = x$.

3. Suppose that the random variable X has the following c.d.f.:

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0, \\ \frac{2}{5}x & \text{for } 0 < x \leq 1, \\ \frac{3}{5}x - \frac{1}{5} & \text{for } 1 < x \leq 2, \\ 1 & \text{for } x > 2. \end{cases}$$

Verify that X has a continuous distribution, and determine the p.d.f. of X.

4. Suppose that the random variable X has a continuous distribution with the following p.d.f.:

$$f(x) = \frac{1}{2}e^{-|x|} \quad \text{for } -\infty < x < \infty.$$

Determine the value x_0 such that $F(x_0) = 0.9$, where $F(x)$ is the c.d.f. of X.

5. Suppose that X_1 and X_2 are i.i.d. random variables, and that each has the uniform distribution on the interval $[0, 1]$. Evaluate $\Pr(X_1^2 + X_2^2 \le 1)$.

6. For each value of $p > 1$, let

$$c(p) = \sum_{x=1}^{\infty} \frac{1}{x^p}.$$

Suppose that the random variable X has a discrete distribution with the following p.f.:

$$f(x) = \frac{1}{c(p)x^p} \quad \text{for } x = 1, 2, \ldots.$$

a. For each fixed positive integer n, determine the probability that X will be divisible by n.

b. Determine the probability that X will be odd.

7. Suppose that X_1 and X_2 are i.i.d. random variables, each of which has the p.f. $f(x)$ specified in Exercise 6. Determine the probability that $X_1 + X_2$ will be even.

8. Suppose that an electronic system comprises four components, and let X_j denote the time until component j fails to operate ($j = 1, 2, 3, 4$). Suppose that X_1, X_2, X_3, and X_4 are i.i.d. random variables, each of which has a continuous distribution with c.d.f. $F(x)$. Suppose that the system will operate as long as both component 1 and at least one of the other three components operate. Determine the c.d.f. of the time until the system fails to operate.

9. Suppose that a box contains a large number of tacks and that the probability X that a particular tack will land with its point up when it is tossed varies from tack to tack in accordance with the following p.d.f.:

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

Suppose that a tack is selected at random from the box and that this tack is then tossed three times independently. Determine the probability that the tack will land with its point up on all three tosses.

10. Suppose that the radius X of a circle is a random variable having the following p.d.f.:

$$f(x) = \begin{cases} \frac{1}{8}(3x+1) & \text{for } 0 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

Determine the p.d.f. of the area of the circle.

11. Suppose that the random variable X has the following p.d.f.:

$$f(x) = \begin{cases} 2e^{-2x} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Construct a random variable $Y = r(X)$ that has the uniform distribution on the interval $[0, 5]$.

12. Suppose that the 12 random variables $X_1, \ldots, X_{12}$ are i.i.d. and each has the uniform distribution on the interval $[0, 20]$. For $j = 0, 1, \ldots, 19$, let I_j denote the interval $(j, j+1)$. Determine the probability that none of the 20 disjoint intervals I_j will contain more than one of the random variables $X_1, \ldots, X_{12}$.

13. Suppose that the joint distribution of X and Y is uniform over a set A in the xy-plane. For which of the following sets A are X and Y independent?

a. A circle with a radius of 1 and with its center at the origin

b. A circle with a radius of 1 and with its center at the point (3, 5)

c. A square with vertices at the four points (1, 1), (1, −1), (−1, −1), and (−1, 1)

d. A rectangle with vertices at the four points (0, 0), (0, 3), (1, 3), and (1, 0)

e. A square with vertices at the four points (0, 0), (1, 1), (0, 2), and (−1, 1)

14. Suppose that X and Y are independent random variables with the following p.d.f.'s:

$$f_1(x) = \begin{cases} 1 & \text{for } 0 < x < 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$f_2(y) = \begin{cases} 8y & \text{for } 0 < y < \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

Determine the value of $\Pr(X > Y)$.

15. Suppose that, on a particular day, two persons A and B arrive at a certain store independently of each other. Suppose that A remains in the store for 15 minutes and B remains in the store for 10 minutes. If the time of arrival of each person has the uniform distribution over the hour between 9:00 A.M. and 10:00 A.M., what is the probability that A and B will be in the store at the same time?

16. Suppose that X and Y have the following joint p.d.f.:

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

Determine **(a)** $\Pr(X < 1/2)$; **(b)** the marginal p.d.f. of X; and **(c)** the conditional p.d.f. of Y given that $X = x$.

Note: More General Distributions. In Example 3.2.7, we introduced a type of distribution that was neither discrete nor continuous. It is possible to define expectations for such distributions also. The definition is rather cumbersome, and we shall not pursue it here.

Summary

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

Exercises

1. Suppose that X has the uniform distribution on the interval $[a, b]$. Find the mean of X.

2. If an integer between 1 and 100 is to be chosen at random, what is the expected value?

3. In a class of 50 students, the number of students n_i of each age i is shown in the following table:

Age i	n_i
18	20
19	22
20	4
21	3
25	1

If a student is to be selected at random from the class, what is the expected value of his age?

4. Suppose that one word is to be selected at random from the sentence THE GIRL PUT ON HER BEAUTIFUL RED HAT. If X denotes the number of letters in the word that is selected, what is the value of $E(X)$?

5. Suppose that one letter is to be selected at random from the 30 letters in the sentence given in Exercise 4. If Y denotes the number of letters in the word in which the selected letter appears, what is the value of $E(Y)$?

6. Suppose that a random variable X has a continuous distribution with the p.d.f. f given in Example 4.1.6. Find the expectation of $1/X$.

7. Suppose that a random variable X has the uniform distribution on the interval $[0, 1]$. Show that the expectation of $1/X$ is infinite.

8. Suppose that X and Y have a continuous joint distribution for which the joint p.d.f. is as follows:

$$f(x, y) = \begin{cases} 12y^2 & \text{for } 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the value of $E(XY)$.

9. Suppose that a point is chosen at random on a stick of unit length and that the stick is broken into two pieces at that point. Find the expected value of the length of the longer piece.

10. Suppose that a particle is released at the origin of the xy-plane and travels into the half-plane where $x > 0$. Suppose that the particle travels in a straight line and that the angle between the positive half of the x-axis and this line is α, which can be either positive or negative. Suppose, finally, that the angle α has the uniform distribution on the interval $[-\pi/2, \pi/2]$. Let Y be the ordinate of the point at which the particle hits the vertical line $x = 1$. Show that the distribution of Y is a Cauchy distribution.

11. Suppose that the random variables $X_1, \dots, X_n$ form a random sample of size n from the uniform distribution on the interval $[0, 1]$. Let $Y_1 = \min\{X_1, \dots, X_n\}$, and let $Y_n = \max\{X_1, \dots, X_n\}$. Find $E(Y_1)$ and $E(Y_n)$.

12. Suppose that the random variables $X_1, \dots, X_n$ form a random sample of size n from a continuous distribution for which the c.d.f. is F, and let the random variables Y_1 and Y_n be defined as in Exercise 11. Find $E[F(Y_1)]$ and $E[F(Y_n)]$.

13. A stock currently sells for \$110 per share. Let the price of the stock at the end of a one-year period be X, which will take one of the values \$100 or \$300. Suppose that you have the option to buy shares of this stock at \$150 per share at the end of that one-year period. Suppose that money

Definition 4.3.2 Interquartile Range (IQR). Let X be a random variable with quantile function $F^{-1}(p)$ for $0 < p < 1$. The *interquartile range (IQR)* is defined to be $F^{-1}(0.75) - F^{-1}(0.25)$.

In words, the IQR is the length of the interval that contains the middle half of the distribution.

Example 4.3.9 The Cauchy Distribution. Let X have the Cauchy distribution. The c.d.f. F of X can be found using a trigonometric substitution in the following integral:

$$F(x) = \int_{-\infty}^{x} \frac{dy}{\pi(1+y^2)} = \frac{1}{2} + \frac{\tan^{-1}(x)}{\pi},$$

where $\tan^{-1}(x)$ is the principal inverse of the tangent function, taking values from $-\pi/2$ to $\pi/2$ as x runs from $-\infty$ to ∞. The quantile function of X is then $F^{-1}(p) = \tan[\pi(p - 1/2)]$ for $0 < p < 1$. The IQR is

$$F^{-1}(0.75) - F^{-1}(0.25) = \tan(\pi/4) - \tan(-\pi/4) = 2.$$

It is not difficult to show that, if $Y = 2X$, then the IQR of Y is 4. (See Exercise 14.) ◀

Summary

The variance of X, denoted by $\text{Var}(X)$, is the mean of $[X - E(X)]^2$ and measures how spread out the distribution of X is. The variance also equals $E(X^2) - [E(X)]^2$. The standard deviation is the square root of the variance. The variance of $aX + b$, where a and b are constants, is $a^2\,\text{Var}(X)$. The variance of the sum of independent random variables is the sum of the variances. As an example, the variance of the binomial distribution with parameters n and p is $np(1-p)$. The interquartile range (IQR) is the difference between the 0.75 and 0.25 quantiles. The IQR is a measure of spread that exists for every distribution.

Exercises

1. Suppose that X has the uniform distribution on the interval $[0, 1]$. Compute the variance of X.

2. Suppose that one word is selected at random from the sentence THE GIRL PUT ON HER BEAUTIFUL RED HAT. If X denotes the number of letters in the word that is selected, what is the value of $\text{Var}(X)$?

3. For all numbers a and b such that $a < b$, find the variance of the uniform distribution on the interval $[a, b]$.

4. Suppose that X is a random variable for which $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$. Show that $E[X(X-1)] = \mu(\mu - 1) + \sigma^2$.

5. Let X be a random variable for which $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$, and let c be an arbitrary constant. Show that

$$E[(X-c)^2] = (\mu - c)^2 + \sigma^2.$$

6. Suppose that X and Y are independent random variables whose variances exist and such that $E(X) = E(Y)$. Show that

$$E[(X-Y)^2] = \text{Var}(X) + \text{Var}(Y).$$

7. Suppose that X and Y are independent random variables for which $\text{Var}(X) = \text{Var}(Y) = 3$. Find the values of **(a)** $\text{Var}(X - Y)$ and **(b)** $\text{Var}(2X - 3Y + 1)$.

8. Construct an example of a distribution for which the mean is finite but the variance is infinite.

9. Let X have the discrete uniform distribution on the integers $1, \ldots, n$. Compute the variance of X. *Hint:* You may wish to use the formula $\sum_{k=1}^{n} k^2 = n(n+1)\cdot(2n+1)/6$.

4.9 Supplementary Exercises

1. Suppose that the random variable X has a continuous distribution with c.d.f. $F(x)$ and p.d.f. f. Suppose also that $E(X)$ exists. Prove that

$$\lim_{x\to\infty} x[1 - F(x)] = 0.$$

Hint: Use the fact that if $E(X)$ exists, then

$$E(X) = \lim_{u\to\infty} \int_{-\infty}^{u} xf(x)\,dx.$$

2. Suppose that the random variable X has a continuous distribution with c.d.f. $F(x)$. Suppose also that $\Pr(X \geq 0) = 1$ and that $E(X)$ exists. Show that

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

Hint: You may use the result proven in Exercise 1.

3. Consider again the conditions of Exercise 2, but suppose now that X has a discrete distribution with c.d.f. $F(x)$, rather than a continuous distribution. Show that the conclusion of Exercise 2 still holds.

4. Suppose that X, Y, and Z are nonnegative random variables such that $\Pr(X + Y + Z \leq 1.3) = 1$. Show that X, Y, and Z cannot possibly have a joint distribution under which each of their marginal distributions is the uniform distribution on the interval [0, 1].

5. Suppose that the random variable X has mean μ and variance σ^2, and that $Y = aX + b$. Determine the values of a and b for which $E(Y) = 0$ and $\text{Var}(Y) = 1$.

6. Determine the expectation of the range of a random sample of size n from the uniform distribution on the interval [0, 1].

7. Suppose that an automobile dealer pays an amount X (in thousands of dollars) for a used car and then sells it for an amount Y. Suppose that the random variables X and Y have the following joint p.d.f.:

$$f(x, y) = \begin{cases} \frac{1}{36}x & \text{for } 0 < x < y < 6, \\ 0 & \text{otherwise.} \end{cases}$$

Determine the dealer's expected gain from the sale.

8. Suppose that $X_1, \dots, X_n$ form a random sample of size n from a continuous distribution with the following p.d.f.:

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

Let $Y_n = \max\{X_1, \dots, X_n\}$. Evaluate $E(Y_n)$.

9. If m is a median of the distribution of X, and if $Y = r(X)$ is either a nondecreasing or a nonincreasing function of X, show that $r(m)$ is a median of the distribution of Y.

10. Suppose that $X_1, \dots, X_n$ are i.i.d. random variables, each of which has a continuous distribution with median m. Let $Y_n = \max\{X_1, \dots, X_n\}$. Determine the value of $\Pr(Y_n > m)$.

11. Suppose that you are going to sell cola at a football game and must decide in advance how much to order. Suppose that the demand for cola at the game, in liters, has a continuous distribution with p.d.f. $f(x)$. Suppose that you make a profit of g cents on each liter that you sell at the game and suffer a loss of c cents on each liter that you order but do not sell. What is the optimal amount of cola for you to order so as to maximize your expected net gain?

12. Suppose that the number of hours X for which a machine will operate before it fails has a continuous distribution with p.d.f. $f(x)$. Suppose that at the time at which the machine begins operating you must decide when you will return to inspect it. If you return before the machine has failed, you incur a cost of b dollars for having wasted an inspection. If you return after the machine has failed, you incur a cost of c dollars per hour for the length of time during which the machine was not operating after its failure. What is the optimal number of hours to wait before you return for inspection in order to minimize your expected cost?

13. Suppose that X and Y are random variables for which $E(X) = 3$, $E(Y) = 1$, $\text{Var}(X) = 4$, and $\text{Var}(Y) = 9$. Let $Z = 5X - Y + 15$. Find $E(Z)$ and $\text{Var}(Z)$ under each of the following conditions: **(a)** X and Y are independent; **(b)** X and Y are uncorrelated; **(c)** the correlation of X and Y is 0.25.

14. Suppose that $X_0, X_1, \dots, X_n$ are independent random variables, each having the same variance σ^2. Let $Y_j = X_j - X_{j-1}$ for $j = 1, \dots, n$, and let $\overline{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$. Determine the value of $\text{Var}(\overline{Y}_n)$.

15. Suppose that $X_1, \dots, X_n$ are random variables for which $\text{Var}(X_i)$ has the same value σ^2 for $i = 1, \dots, n$ and $\rho(X_i, X_j)$ has the same value ρ for every pair of values i and j such that $i \neq j$. Prove that $\rho \geq -\frac{1}{n-1}$.

16. Suppose that the joint distribution of X and Y is the uniform distribution over a rectangle with sides parallel to the coordinate axes in the xy-plane. Determine the correlation of X and Y.

17. Suppose that n letters are put at random into n envelopes, as in the matching problem described in Sec. 1.10. Determine the variance of the number of letters that are placed in the correct envelopes.

5.11 Supplementary Exercises

1. Let X and P be random variables. Suppose that the conditional distribution of X given $P = p$ is the binomial distribution with parameters n and p. Suppose that the distribution of P is the beta distribution with parameters $\alpha = 1$ and $\beta = 1$. Find the marginal distribution of X.

2. Suppose that X, Y, and Z are i.i.d. random variables and each has the standard normal distribution. Evaluate $\Pr(3X + 2Y < 6Z - 7)$.

3. Suppose that X and Y are independent Poisson random variables such that $\text{Var}(X) + \text{Var}(Y) = 5$. Evaluate $\Pr(X + Y < 2)$.

4. Suppose that X has a normal distribution such that $\Pr(X < 116) = 0.20$ and $\Pr(X < 328) = 0.90$. Determine the mean and the variance of X.

5. Suppose that a random sample of four observations is drawn from the Poisson distribution with mean λ, and let $\overline{X}$ denote the sample mean. Show that

$$\Pr\left(\overline{X} < \frac{1}{2}\right) = (4\lambda + 1)e^{-4\lambda}.$$

6. The lifetime X of an electronic component has the exponential distribution such that $\Pr(X \le 1000) = 0.75$. What is the expected lifetime of the component?

7. Suppose that X has the normal distribution with mean μ and variance σ^2. Express $E(X^3)$ in terms of μ and σ^2.

8. Suppose that a random sample of 16 observations is drawn from the normal distribution with mean μ and standard deviation 12, and that independently another random sample of 25 observations is drawn from the normal distribution with the same mean μ and standard deviation 20. Let $\overline{X}$ and $\overline{Y}$ denote the sample means of the two samples. Evaluate $\Pr(|\overline{X} - \overline{Y}| < 5)$.

9. Suppose that men arrive at a ticket counter according to a Poisson process at the rate of 120 per hour, and women arrive according to an independent Poisson process at the rate of 60 per hour. Determine the probability that four or fewer people arrive in a one-minute period.

10. Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables, each of which has m.g.f. $\psi(t)$. Let $Y = X_1 + \cdots + X_N$, where the number of terms N in this sum is a random variable having the Poisson distribution with mean λ. Assume that N and $X_1, X_2, \ldots$ are independent, and $Y = 0$ if $N = 0$. Determine the m.g.f. of Y.

11. Every Sunday morning, two children, Craig and Jill, independently try to launch their model airplanes. On each Sunday, Craig has probability 1/3 of a successful launch, and Jill has probability 1/5 of a successful launch. Determine the expected number of Sundays required until at least one of the two children has a successful launch.

12. Suppose that a fair coin is tossed until at least one head and at least one tail have been obtained. Let X denote the number of tosses that are required. Find the p.f. of X.

13. Suppose that a pair of balanced dice are rolled 120 times, and let X denote the number of rolls on which the sum of the two numbers is 12. Use the Poisson approximation to approximate $\Pr(X = 3)$.

14. Suppose that $X_1, \ldots, X_n$ form a random sample from the uniform distribution on the interval $[0, 1]$. Let $Y_1 = \min\{X_1, \ldots, X_n\}$, $Y_n = \max\{X_1, \ldots, X_n\}$, and $W = Y_n - Y_1$. Show that each of the random variables Y_1, Y_n, and W has a beta distribution.

15. Suppose that events occur in accordance with a Poisson process at the rate of five events per hour.

a. Determine the distribution of the waiting time T_1 until the first event occurs.

b. Determine the distribution of the total waiting time T_k until k events have occurred.

c. Determine the probability that none of the first k events will occur within 20 minutes of one another.

16. Suppose that five components are functioning simultaneously, that the lifetimes of the components are i.i.d., and that each lifetime has the exponential distribution with parameter β. Let T_1 denote the time from the beginning of the process until one of the components fails; and let T_5 denote the total time until all five components have failed. Evaluate $\text{Cov}(T_1, T_5)$.

17. Suppose that X_1 and X_2 are independent random variables, and X_i has the exponential distribution with parameter β_i $(i = 1, 2)$. Show that for each constant $k > 0$,

$$\Pr(X_1 > kX_2) = \frac{\beta_2}{k\beta_1 + \beta_2}.$$

18. Suppose that 15,000 people in a city with a population of 500,000 are watching a certain television program. If 200 people in the city are contacted at random, what is the approximate probability that fewer than four of them are watching the program?

19. Suppose that it is desired to estimate the proportion of persons in a large population who have a certain characteristic. A random sample of 100 persons is selected from the population without replacement, and the proportion $\overline{X}$ of persons in the sample who have the characteristic is observed. Show that, no matter how large the population is, the standard deviation of $\overline{X}$ is at most 0.05.

20. Suppose that X has the binomial distribution with parameters n and p, and that Y has the negative binomial distribution with parameters r and p, where r is a positive integer. Show that $\Pr(X < r) = \Pr(Y > n - r)$ by showing that both the left side and the right side of this equation can be regarded as the probability of the same event in a sequence of Bernoulli trials with probability p of success.

21. Suppose that X has the Poisson distribution with mean λt, and that Y has the gamma distribution with parameters $\alpha = k$ and $\beta = \lambda$, where k is a positive integer. Show that $\Pr(X \geq k) = \Pr(Y \leq t)$ by showing that both the left side and the right side of this equation can be regarded as the probability of the same event in a Poisson process in which the expected number of occurrences per unit of time is λ.

22. Suppose that X is a random variable having a continuous distribution with p.d.f. $f(x)$ and c.d.f. $F(x)$, and for which $\Pr(X > 0) = 1$. Let the failure rate $h(x)$ be as defined in Exercise 18 of Sec. 5.7. Show that

$$\exp\left[-\int_0^x h(t)\,dt\right] = 1 - F(x).$$

23. Suppose that 40 percent of the students in a large population are freshmen, 30 percent are sophomores, 20 percent are juniors, and 10 percent are seniors. Suppose that 10 students are selected at random from the population, and let X_1, X_2, X_3, X_4 denote, respectively, the numbers of freshmen, sophomores, juniors, and seniors that are obtained.

a. Determine $\rho(X_i, X_j)$ for each pair of values i and j $(i < j)$.

b. For what values of i and j $(i < j)$ is $\rho(X_i, X_j)$ most negative?

c. For what values of i and j $(i < j)$ is $\rho(X_i, X_j)$ closest to 0?

24. Suppose that X_1 and X_2 have the bivariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ. Determine the distribution of $X_1 - 3X_2$.

25. Suppose that X has the standard normal distribution, and the conditional distribution of Y given X is the normal distribution with mean $2X - 3$ and variance 12. Determine the marginal distribution of Y and the value of $\rho(X, Y)$.

26. Suppose that X_1 and X_2 have a bivariate normal distribution with $E(X_2) = 0$. Evaluate $E(X_1^2 X_2)$.