f) Identify the probability distribution of X as right-skewed, symmetric, or left-skewed without consulting its probability distribution or drawing its probability histogram.
g) Draw a probability histogram for X.

5.43 Sickle cell anemia is an inherited blood disease that occurs primarily in blacks. In the United States, about 15 of every 10,000 black children have sickle cell anemia. The red blood cells of an affected person are abnormal; the result is severe chronic anemia (inability to carry the required amount of oxygen), which causes headaches, shortness of breath, jaundice, increased risk of pneumococcal pneumonia and gallstones, and other severe problems. Sickle cell anemia occurs in children who inherit an abnormal type of hemoglobin, called hemoglobin S, from both parents. If hemoglobin S is inherited from only one parent, the person is said to have sickle cell trait and is generally free from symptoms. There is a 50% chance that a person who has sickle cell trait will pass hemoglobin S to an offspring.

a) Obtain the probability that a child of two people who have sickle cell trait will have sickle cell anemia.
b) If two people who have sickle cell trait have five children, determine the probability that at least one of the children will have sickle cell anemia.
c) If two people who have sickle cell trait have five children, find the probability distribution of the number of those children who will have sickle cell anemia.
d) Construct a probability histogram for the probability distribution in part (c).

5.44 If all sex distributions are equally likely, what proportion of families with five children have three girls and two boys?

5.45 A baseball player has a batting average of .260. Suppose that you observe successive at-bats of the player and note for each at-bat whether the player gets a hit.

a) Under what conditions is the assumption of Bernoulli trials appropriate?
b) Assuming your conditions in part (a), what is the probability that the player will get two or more hits in his next four times at-bat?

5.46 Sixty percent of all voters in a state intend to vote "yes" in a referendum. An opinion poll took a sample of 20 voters. No precaution was taken against the unlikely event of choosing the same voter more than once. Find the probability that in the sample of 20 there are more voters who intend to vote "yes" than voters who don't intend to vote "yes."

5.47 Consider 15 Bernoulli trials with success probability p.

a) Assuming that $p = 0.5$, what is the probability that the number of successes is between six and nine, inclusive? What is the most likely number of successes?
b) Repeat part (a) for $p = 0.4$.

5.48 Availability of statistical software or binomial tables is useful for this problem. Suppose that you throw a balanced die 100 times.

a) What is the probability that you get a 1 at most 4 times? What would you think if that happened?
b) What is the probability that you get a 1 more than 32 times? What would you think if that happened?

5.49 Simultaneously and independently, each of n people make a single toss of a coin with probability p of a head. What is the probability of an "odd man"—that is, one person gets a different result from all the other people?

5.50 How long must a sequence of random decimal digits be so that the probability of getting a 6 or a 7 will exceed 0.95?

5.61 Five cards are selected at random without replacement from an ordinary deck of 52 playing cards.

a) What is the probability that exactly 3 face cards are obtained?

b) Identify and provide a formula for the probability distribution of the number of face cards obtained.

5.62 Estimating a population proportion: Many statistical studies are concerned with the proportion of members of a finite population that have a specified attribute, called the *population proportion.* In practice, we mostly rely on sampling and use the sample data to estimate the population proportion. Suppose that a random sample of size n is taken without replacement from a population of size N in which the proportion of members having the specified attribute is p. Intuitively, it makes sense to estimate the population proportion, p, by the *sample proportion,* $\hat{p} = X/n$, where X denotes the number of members sampled that have the specified attribute. Determine the PMF of the random variable $\hat{p}$.

5.63 From 10 pills, of which 5 are placebos, you are to randomly select and take 5 as part of an experiment on the effectiveness of a new treatment. Find the probability that

a) you select at least 2 placebos.

b) the first three pills you select are placebos.

5.64 As reported by Television Bureau of Advertising, Inc., in *Trends in Television,* 84.2% of U.S. households have a VCR. If six U.S. households are randomly selected without replacement, what is the (approximate) probability that the number of households sampled that have a VCR will be

a) exactly four? **b)** at least four? **c)** at most four?

d) between two and five, inclusive?

e) Determine a formula that approximates the PMF of the random variable Y, the number of households of the six sampled that have a VCR.

f) Strictly speaking, why is the PMF that you obtained in part (e) only approximately correct?

g) Determine a formula that provides the exact PMF of the random variable Y in terms of the number, N, of U.S. households.

5.65 Bin I contains 20 parts, of which 5 are defective. Bin II contains 15 parts, of which 4 are defective. One of these two bins is chosen at random and 3 parts are randomly selected from the bin chosen. If 2 of the 3 parts obtained are defective, what is the probability that Bin I was chosen?

5.66 An upper-level probability class has six undergraduate students and four graduate students. A random sample of three students is taken from the class. Let X denote the number of undergraduate students selected. Identify, obtain a formula for, and tabulate the PMF of the random variable X if the sampling is

a) without replacement. **b)** with replacement.

c) Compare your results in parts (a) and (b).

5.67 Refer to Example 5.14 on page 208, where X denotes the number of defective TVs obtained when 5 TVs are randomly selected without replacement from a lot of 100 TVs in which 6 are defective.

a) Construct a table for the PMF of the random variable X similar to Table 5.14 on page 213. Round each probability to eight decimal places.

b) Approximate the PMF by the appropriate binomial distribution. Construct a table similar to Table 5.15 on page 215.

c) Comment on the accuracy here of the binomial approximation to the hypergeometric distribution.

5.68 A dictionary contains 80,000 words, of which a certain student knows 30,000. One hundred of the 80,000 words in the dictionary are chosen at random without replacement.
a) Determine the probability that the student knows exactly 40 of the 100 words obtained.
b) Determine the probability that the student knows at least 35, but no more than 40, of the 100 words obtained.
c) Would a binomial approximation to the hypergeometric distribution be justified in this case? Explain your answer.
d) Another student knows 38 of the 100 words obtained. Estimate the size of that student's vocabulary, assuming that the student knows very few words that aren't in this particular dictionary.

Theory Exercises

5.69 Prove Proposition 5.5 on page 210—that is, derive Equation (5.12) which gives the PMF of a hypergeometric random variable with parameters N, n, and p.

5.70 Prove Proposition 5.6 on page 215, the binomial approximation to the hypergeometric distribution.

5.71 Consider a finite population of size N in which each member is classified as either having or not having a specified attribute; use p to denote the proportion of the population having the specified attribute. Suppose that a random sample of size n is taken from the population and let X denote the number of members sampled that have the specified attribute. Set $q = 1 - p$.
a) If the sampling is with replacement, prove that

$$p_X(x) = \binom{n}{x} \frac{(Np)^x (Nq)^{n-x}}{(N)^n}, \qquad x = 0, 1, \ldots, n,$$

and $p_X(x) = 0$ otherwise.
b) Verify that the PMF in part (a) is that of a binomial distribution with parameters n and p. Thus part (a) gives an alternative form for a binomial PMF in the case of sampling with replacement from a finite population.
c) If the sampling is without replacement, prove that

$$p_X(x) = \binom{n}{x} \frac{(Np)_x (Nq)_{n-x}}{(N)_n}, \qquad x = 0, 1, \ldots, n,$$

and $p_X(x) = 0$ otherwise.
d) Verify that the PMF in part (c) is that of a hypergeometric distribution with parameters N, n, and p. Thus part (c) gives an alternative form for a hypergeometric PMF.

5.72 Let N, M, and n be positive integers with $M \le N$ and $n \le N$.
a) Prove that $\sum_{k=0}^{n} \binom{M}{k}\binom{N-M}{n-k} / \binom{N}{n} = 1$ by applying Vandermonde's identity, presented in Exercise 3.48 on page 109.
b) Establish the identity in part (a) without using Vandermonde's identity or doing any computations, but rather by using a probabilistic argument.
c) Use part (b) to establish Vandermonde's identity.

Advanced Exercises

5.73 An office has 20 employees, 12 men and 8 women. A committee of size 5 that was chosen to organize the office Christmas party has 4 women and 1 man. Do the men have reason to complain that the committee was not randomly selected? Explain your answer.

Let's apply Proposition 5.8 to the Poisson distribution shown in Figure 5.6. In this case, $\lambda = 6.9$ and, so, according to Proposition 5.8, the probabilities increase until $x = \lfloor 6.9 \rfloor = 6$ and then decrease thereafter. This fact is borne out by both Table 5.17 and Figure 5.6.

5.5 Basic Exercises

5.77 Suppose that X has the Poisson distribution with parameter $\lambda = 3$. Determine
a) $P(X = 3)$. **b)** $P(X < 3)$. **c)** $P(X > 3)$. **d)** $P(X \leq 3)$. **e)** $P(X \geq 3)$.

5.78 A paper by L. F. Richardson, published in the *Journal of the Royal Statistical Society*, analyzed the distribution of wars over time. The data indicate that the number of wars that begin during a given calendar year has approximately the Poisson distribution with parameter $\lambda = 0.7$. If a calendar year is selected at random, find the probability that the number of wars that begin during that calendar year will be
a) zero. **b)** at most two. **c)** between one and three, inclusive.

5.79 M. F. Driscoll and N. A. Weiss discussed the modeling of motel reservation networks in "An Application of Queuing Theory to Reservation Networks" (*TIMS*, 1976, Vol. 22, pp. 540–546). They defined a Type 1 call to be a call from a motel's computer terminal to the national reservation center. For a certain motel, the number of Type 1 calls per hour has a Poisson distribution with parameter $\lambda = 1.7$. Find the probability that the number of Type 1 calls made from this motel during a period of 1 hour will be
a) exactly one. **b)** at most two. **c)** at least two.

Let X denote the number of Type 1 calls made by the motel during a 1-hour period.
d) Construct a table of probabilities for the random variable X. Compute the probabilities until they are zero to three decimal places.
e) Draw a histogram of the probabilities in part (d).

5.80 The second leading genetic cause of mental retardation is Fragile X Syndrome, named for the fragile appearance of the tip of the X chromosome in affected individuals. One in 1500 males are affected worldwide, with no ethnic bias. For a sample of 10,000 males, use the Poisson approximation to the binomial distribution to determine the probability that the number who have Fragile X Syndrome
a) exceeds 7. **b)** is at most 10.
c) What is the maximum possible error that you made in parts (a) and (b) by using the Poisson approximation to the binomial distribution?

5.81 A most amazing event occurred during the second round of the 1989 U.S. Open at Oak Hill in Pittsford, New York. Four golfers—Doug Weaver, Mark Wiebe, Jerry Pate, and Nick Price—made holes in one on the sixth hole. According to the experts, the odds against a PGA golfer making a hole in one are 3708 to 1—that is, the probability of making a hole in one is 1/3709. Determine the probability to nine decimal places that at least four of the 155 golfers playing the second round would get a hole in one on the sixth hole by using
a) the binomial distribution. **b)** the Poisson approximation to the binomial.
c) What assumptions are you making in obtaining your answers in parts (a) and (b)? Do you think that those assumptions are reasonable? Explain your reasoning.

5.82 At a service counter, arrivals occur at an average rate of 20 per hour and follow a Poisson distribution. Let X denote the number of arrivals during a particular hour.

a) Using statistical software, we found that $P(X \leq 20) = 0.559$. Based on that result, find $P(X \geq 20)$.

b) If there were at least 20 arrivals during a particular hour, what is the probability that there were exactly 25 arrivals?

5.83 In a certain population, the number of colds a person gets in a year has a $\mathcal{P}(3)$ distribution. A new anti-cold drug lowers λ from 3 to 0.75 and is effective for 8 out of 10 people. At the beginning of last year, the entire population was given the drug. At the end of last year, one person was selected at random from the population and was found to have had only one cold during the year. What is the probability that the drug was effective for this person?

5.84 Characteristic α occurs in about 0.5% of a population. A random sample is to be taken without replacement from the population. How large must the sample size be to ensure that the chances exceed 90% of the sample containing a person with characteristic α? Answer this question by using both the exact distribution and the Poisson approximation. Compare your answers.

5.85 The number of eggs that the female grunge beetle lays in her nest has a $\mathcal{P}(4)$ distribution. Assuming that an empty nest cannot be recognized, determine the PMF of the number of eggs observed in a nest.

Theory Exercises

5.86 Poisson approximation to the binomial: Let λ be a positive real number. Suppose that, for each $n \in \mathcal{N}$, the random variable X_n has the binomial distribution with parameters n and p_n, where $np_n \to \lambda$ as $n \to \infty$.

a) Prove that

$$\lim_{n\to\infty} p_{X_n}(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots.$$

Hint: Use the fact that, if $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers that converges to a, then $(1 + a_n/n)^n \to e^a$ as $n \to \infty$.

b) Compare the result in part (a) with Proposition 5.7 on page 220.

5.87 Prove Proposition 5.8 on page 225.

5.88 Let k be a nonnegative integer and let a be a nonnegative real number.

a) Use integration by parts to verify the identity

$$\frac{1}{(k+1)!}\int_a^{\infty} t^{k+1}e^{-t}\,dt = e^{-a}\frac{a^{k+1}}{(k+1)!} + \frac{1}{k!}\int_a^{\infty} t^k e^{-t}\,dt.$$

b) Use part (a) and mathematical induction to prove that $\int_0^{\infty} t^k e^{-t}\,dt = k!$.

c) Let $X \sim \mathcal{P}(\lambda)$. Prove that

$$P(X \leq k) = \frac{1}{k!}\int_{\lambda}^{\infty} t^k e^{-t}\,dt = \frac{1}{k!}\Gamma(k+1, \lambda),$$

where $\Gamma(\alpha, x) = \int_x^{\infty} t^{\alpha-1}e^{-t}\,dt$ is called an *incomplete gamma function*.

d) Let $X \sim \mathcal{P}(\lambda)$. Prove that

$$P(X > k) = \frac{1}{k!}\int_0^{\lambda} t^k e^{-t}\,dt = \frac{1}{k!}\gamma(k+1, \lambda),$$

where $\gamma(\alpha, x) = \int_0^x t^{\alpha-1}e^{-t}\,dt$ is also called an *incomplete gamma function*.

trials. The associated PMF is

$$p_X(x) = p(1-p)^{x-1}, \qquad x = 1, 2, \ldots, \tag{5.37}$$

and $p_X(x) = 0$ otherwise. Here, of course, p denotes the success probability.

Instead of considering the number of trials, X, up to and including the first success, we can consider the number of trials, Y, before the first success or, equivalently, the number of failures before the first success. We can determine the PMF of Y directly, or we can use Equation (5.37) and the fact that $Y = X - 1$. In any case, we find that the PMF of the random variable Y is

$$p_Y(y) = p(1-p)^y, \qquad y = 0, 1, \ldots, \tag{5.38}$$

and $p_Y(y) = 0$ otherwise.

Some researchers and textbook authors allude to the geometric distribution as that given by Equation (5.38) instead of Equation (5.37). That is, they define a geometric random variable as the number of failures before the first success rather than the number of trials until the first success. Be sure that you know which form is being considered when you see a reference to a geometric distribution or a geometric random variable.

s 5.6 Basic Exercises

5.93 According to the *Daily Racing Form,* the probability is about 0.67 that the favorite in a horse race will finish in the money (first, second, or third place). Suppose that you always bet the favorite "across the board," which means that you win something if the favorite finishes in the money. Let X denote the number of races that you bet until you win something.

a) Determine and identify the PMF of the random variable X.

b) Find the probability that the number of races that you bet until you win something is exactly three; at least three; at most three.

c) How many races must you bet to be at least 99% sure of winning something?

5.94 A baseball player has a batting average of .260. Suppose that you observe successive at-bats of the player and note for each at-bat whether the player gets a hit. Presuming that the assumption of Bernoulli trials is appropriate, what is the probability that the first hit by the player occurs

a) on his fifth at-bat?

b) after his fifth at-bat?

c) between his third and tenth at-bats, inclusive?

5.95 Let $X \sim \mathcal{G}(p)$. Determine

a) $P(X \text{ is even})$. **b)** $P(X \text{ is odd})$. **c)** $P(2 \le X \le 9 \mid X \ge 4)$.

d) $P(X = k \mid X \le n)$ for $k = 1, 2, \ldots, n$.

e) $P(X = n - k \mid X < n)$ for $k = 1, 2, \ldots, n-1$.

5.96 Let X be a positive-integer valued random variable. If X has the lack-of-memory property, which of these two equations must hold:

$$P(X = 10 \mid X > 6) = P(X = 4) \qquad \text{or} \qquad P(X = 10 \mid X > 6) = P(X = 10)?$$

Explain.

5.97 Let X have the geometric distribution with parameter p. For k and n positive integers, with $k > n$, determine $P(X = k \mid X > n)$
a) without using the lack-of-memory property.
b) by using the lack-of-memory property.

5.98 Coins I and II have probabilities p_1 and p_2 of a head, respectively. One of these two coins is selected at random and tossed until the first tail occurs. Let X denote the number of tosses required.
a) Determine the PMF of the random variable X.
b) Determine a sufficient condition on p_1 and p_2 that makes X a geometric random variable. In that case, what is the parameter of the geometric distribution for X?

5.99 Consider a finite population of size N in which the proportion of the population that has a specified attribute is p. Suppose that members of the population are randomly sampled one at a time until a member with the specified attribute is obtained. Let X denote the number of members sampled. Determine the PMF of the random variable X if the sampling is
a) with replacement.
b) without replacement.
c) Use the result from part (b) to solve Exercise 4.28 on page 144.

5.100 Tom has a coin with probability p_1 of a head and Dick has a coin with probability p_2 of a head. Tom and Dick alternate tossing their coins, with Tom going first. The first person to get a head wins.
a) Find the probability that Tom wins. *Hint:* First obtain the probability that Tom wins on toss $2n + 1$.
b) Evaluate the probability in part (a) if both coins are balanced.
c) What must be the relationship between p_1 and p_2 to make the game fair?
d) If Dick's coin is balanced, what must be the probability of a head for Tom's coin to make the game fair?

5.101 Refer to Exercise 5.100, but assume the person who goes first is chosen at random.
a) Find the probability that Tom wins.
b) What must be the relationship between p_1 and p_2 to make the game fair?

5.102 Refer to Exercise 5.85 on page 227. Let Y denote the number of nests examined until a nest containing six eggs is observed. Identify the PMF of the random variable Y.

Theory Exercises

5.103 Verify that Equation (5.29) on page 229 really does define a probability mass function.

5.104 Proposition 4.6 on page 155 states that, if E is an event with positive probability and the random experiment is repeated independently and indefinitely, the probability is 1 that event E will eventually occur.
a) Use the forementioned result to deduce that, in Bernoulli trials, the probability is 1 that a success will eventually occur.
b) Use the fact that a geometric PMF really does define a probability mass function to obtain the result of part (a).

5.105 Verify Equation (5.31) on page 230—that is, show that, for any random variable X,

$$P(a < X \leq b) = P(X > a) - P(X > b),$$

for all real numbers a and b with $a < b$.

5.112 Suppose that $A_1, A_2, \ldots$ is a countable collection of mutually exclusive events of a sample space. Show that $I_{\bigcup_n A_n} = \sum_n I_{A_n}$.

5.113 Suppose that X has the binomial distribution with parameters n and p.
a) Express X as the sum of n indicator random variables. Interpret your answer.
b) Express X as the sum of n Bernoulli random variables.

5.114 Let $\Omega = \{\omega_1, \ldots, \omega_N\}$ be a finite sample space with equally likely outcomes (i.e., a classical probability model). Define the random variable X on Ω by $X(\omega_k) = k$ for $k = 1, 2, \ldots, N$. Determine and identify the PMF of X.

5.115 Let S consist of the 10 decimal digits. Suppose that a number X is chosen according to the discrete uniform distribution on S and then a number Y is chosen according to the discrete uniform distribution on S with X removed. Determine and identify the PMF of the random variable Y
a) by conditioning on the value of X.
b) by using a symmetry argument.

5.116 According to the *Daily Racing Form,* the probability is about 0.67 that the favorite in a horse race will finish in the money (first, second, or third place). Suppose that you always bet the favorite "across the board," which means that you win something if the favorite finishes in the money. Let X denote the number of races that you bet until you win something three times.
a) Determine and identify the PMF of the random variable X.
b) Find the probability that the number of races that you bet until you win something three times is exactly four; at least four; at most four.

5.117 A baseball player has a batting average of .260. Suppose that you observe successive at-bats of the player and note for each at-bat whether the player gets a hit. Presuming that the assumption of Bernoulli trials is appropriate, what is the probability that the second hit by the player occurs
a) on his fifth at-bat?
b) after his fifth at-bat?
c) between his third and tenth at-bats, inclusive?

5.118 Let X have the negative binomial distribution with parameters r and p. For what values of r does X have the lack-of-memory property? Explain your answer.

5.119 Two balanced dice are rolled until the fourth time a sum of 7 or 11 occurs. What is the probability that it will take more than six rolls?

5.120 Suppose that $X \sim \mathcal{NB}(r, p)$ and that $Y \sim \mathcal{B}(n, p)$.
a) Give a probabilistic argument to show that $P(X > n) = P(Y < r)$.
b) Use the FPF to express the equality in part (a) in terms of PMFs.
c) Using the complementation rule, how many terms of the PMF of X must be evaluated to determine $P(X > n)$?
d) How many terms of the PMF of Y must be evaluated to determine $P(Y < r)$?
e) Use your answers from parts (c) and (d) to comment on the computational savings of using the result of part (a) to evaluate $P(X > n)$ when n is large relative to r.

5.121 Verify that the PMF of a negative binomial random variable with parameters r and p can be expressed in the form given in Equation (5.44) on page 241.

5.122 Refer to Example 5.26 on page 242. Find the probability that the telemarketer makes
a) the second sale by the fifth call and the fifth sale on the fifteenth call.
b) the second sale by the fifth call and the fifth sale by the fifteenth call.