Wackerly, Mendenhall, Scheaffer

Discrete
Random Variables:
Binomial, Geometric
Hypergeometric, Poisson
Problems

questioned prior to a local election either favors candidate Jones or does not. In this section we are concerned with experiments, known as *binomial experiments*, that exhibit the following characteristics.

DEFINITION 3.6

A *binomial experiment* possesses the following properties:

1. The experiment consists of a fixed number, n, of identical trials.
2. Each trial results in one of two outcomes: success, S, or failure, F.
3. The probability of success on a single trial is equal to some value p and remains the same from trial to trial. The probability of a failure is equal to $q = (1 - p)$.
4. The trials are independent.
5. The random variable of interest is Y, the number of successes observed during the n trials.

Determining whether a particular experiment is a binomial experiment requires examining the experiment for each of the characteristics just listed. Notice that the random variable of interest is the number of successes observed in the n trials. It is important to realize that a success is not necessarily "good" in the everyday sense of the word. In our discussions, success is merely a name for one of the two possible outcomes on a single trial of an experiment.

EXAMPLE 3.5 An early-warning detection system for aircraft consists of four identical radar units operating independently of one another. Suppose that each has a probability of .95 of detecting an intruding aircraft. When an intruding aircraft enters the scene, the random variable of interest is Y, the number of radar units that *do not detect* the plane. Is this a binomial experiment?

Solution To decide whether this is a binomial experiment, we must determine whether each of the five requirements in Definition 3.6 is met. Notice that the random variable of interest is Y, the number of radar units that do *not* detect an aircraft. The random variable of interest in a binomial experiment is always the number of successes; consequently, the present experiment can be binomial only if we call the event *do not detect* a success. We now examine the experiment for the five characteristics of the binomial experiment.

1. The experiment involves four identical trials. Each trial consists of determining whether (or not) a particular radar unit detects the aircraft.
2. Each trial results in one of two outcomes. Because the random variable of interest is the number of successes, S denotes that the aircraft was not detected, and F denotes that it was detected.
3. Because all the radar units detect aircraft with equal probability, the probability of an S on each trial is the same, and $p = P(S) = P(\text{do not detect}) = .05$.

4. The trials are independent because the units operate independently.
5. The random variable of interest is Y, the number of successes in four trials.

Thus, the experiment is a binomial experiment, with $n = 4$, $p = .05$, and $q = 1 - .05 = .95$. ■

EXAMPLE 3.6 Suppose that 40% of a large population of registered voters favor candidate Jones. A random sample of $n = 10$ voters will be selected, and Y, the number favoring Jones, is to be observed. Does this experiment meet the requirements of a binomial experiment?

Solution If each of the ten people is selected at random from the population, then we have ten nearly identical trials, with each trial resulting in a person either favoring Jones (S) or not favoring Jones (F). The random variable of interest is then the number of successes in the ten trials. For the first person selected, the probability of favoring Jones (S) is .4. But what can be said about the *unconditional* probability that the second person will favor Jones? In Exercise 3.35 you will show that *unconditionally* the probability that the second person favors Jones is also .4. Thus, the probability of a success S stays the same from trial to trial. However, the *conditional* probability of a success on later trials depends on the number of successes in the previous trials. If the population of voters is large, removal of one person will not substantially change the fraction of voters favoring Jones, and the *conditional* probability that the second person favors Jones will be very close to .4. In general, if the population is large and the sample size is relatively small, the *conditional* probability of success on a later trial given the number of successes on the previous trials will stay approximately the same regardless of the outcomes on previous trials. Thus, the trials will be approximately independent and so sampling problems of this type are approximately binomial. ■

If the sample size in Example 3.6 was large relative to the population size (say, 10% of the population), the *conditional* probability of selecting a supporter of Jones on a later selection would be significantly altered by the preferences of persons selected earlier in the experiment, and the experiment would not be binomial. The hypergeometric probability distribution, the topic of Section 3.7, is the appropriate probability model to be used when the sample size is large relative to the population size.

You may wish to refine your ability to identify binomial experiments by reexamining the exercises at the end of Chapter 2. Several of the experiments in those exercises are binomial or approximately binomial experiments.

The binomial probability distribution $p(y)$ can be derived by applying the sample-point approach to find the probability that the experiment yields y successes. Each sample point in the sample space can be characterized by an n-tuple involving the

the proportion of "successes" in our sample, in this case 6/20. In the next section, we will apply this same technique to obtain an estimate that is not initially so intuitive. As we will see in Chapter 9, the estimate that we just obtained is the *maximum likelihood* estimate for p and the procedure used above is an example of the application of the *method of maximum likelihood.* ■

Problems on Binomial Random Variables

Exercises

3.35 Consider the population of voters described in Example 3.6. Suppose that there are $N = 5000$ voters in the population, 40% of whom favor Jones. Identify the event *favors Jones* as a success S. It is evident that the probability of S on trial 1 is .40. Consider the event B that S occurs on the second trial. Then B can occur two ways: The first two trials are both successes *or* the first trial is a failure and the second is a success. Show that $P(B) = .4$. What is $P(B|$ the first trial is $S)$? Does this *conditional* probability differ markedly from $P(B)$?

3.36 **a** A meteorologist in Denver recorded $Y =$ the number of days of rain during a 30-day period. Does Y have a binomial distribution? If so, are the values of both n and p given?

b A market research firm has hired operators who conduct telephone surveys. A computer is used to randomly dial a telephone number, and the operator asks the answering person whether she has time to answer some questions. Let $Y =$ the number of calls made until the first person replies that she is willing to answer the questions. Is this a binomial experiment? Explain.

3.37 In 2003, the average combined SAT score (math and verbal) for college-bound students in the United States was 1026. Suppose that approximately 45% of all high school graduates took this test and that 100 high school graduates are randomly selected from among all high school grads in the United States. Which of the following random variables has a distribution that can be approximated by a binomial distribution? Whenever possible, give the values for n and p.

a The number of students who took the SAT

b The scores of the 100 students in the sample

c The number of students in the sample who scored above average on the SAT

d The amount of time required by each student to complete the SAT

e The number of female high school grads in the sample

3.38 The manufacturer of a low-calorie dairy drink wishes to compare the taste appeal of a new formula (formula B) with that of the standard formula (formula A). Each of four judges is given three glasses in random order, two containing formula A and the other containing formula B. Each judge is asked to state which glass he or she most enjoyed. Suppose that the two formulas are equally attractive. Let Y be the number of judges stating a preference for the new formula.

a Find the probability function for Y.

b What is the probability that at least three of the four judges state a preference for the new formula?

c Find the expected value of Y.

d Find the variance of Y.

3.39 A complex electronic system is built with a certain number of backup components in its subsystems. One subsystem has four identical components, each with a probability of .2 of failing in less than 1000 hours. The subsystem will operate if any two of the four components are operating. Assume that the components operate independently. Find the probability that

a exactly two of the four components last longer than 1000 hours.

b the subsystem operates longer than 1000 hours.

3.40 The probability that a patient recovers from a stomach disease is .8. Suppose 20 people are known to have contracted this disease. What is the probability that

a exactly 14 recover?

b at least 10 recover?

c at least 14 but not more than 18 recover?

d at most 16 recover?

3.41 A multiple-choice examination has 15 questions, each with five possible answers, only one of which is correct. Suppose that one of the students who takes the examination answers each of the questions with an independent random guess. What is the probability that he answers at least ten questions correctly?

3.42 Refer to Exercise 3.41. What is the probability that a student answers at least ten questions correctly if

a for each question, the student can correctly eliminate one of the wrong answers and subsequently answers each of the questions with an independent random guess among the remaining answers?

b he can correctly eliminate two wrong answers for each question and randomly chooses from among the remaining answers?

3.43 Many utility companies promote energy conservation by offering discount rates to consumers who keep their energy usage below certain established subsidy standards. A recent EPA report notes that 70% of the island residents of Puerto Rico have reduced their electricity usage sufficiently to qualify for discounted rates. If five residential subscribers are randomly selected from San Juan, Puerto Rico, find the probability of each of the following events:

a All five qualify for the favorable rates.

b At least four qualify for the favorable rates.

3.44 A new surgical procedure is successful with a probability of p. Assume that the operation is performed five times and the results are independent of one another. What is the probability that

a all five operations are successful if $p = .8$?

b exactly four are successful if $p = .6$?

c less than two are successful if $p = .3$?

3.45 A fire-detection device utilizes three temperature-sensitive cells acting independently of each other in such a manner that any one or more may activate the alarm. Each cell possesses a probability of $p = .8$ of activating the alarm when the temperature reaches 100° Celsius or more. Let Y equal the number of cells activating the alarm when the temperature reaches 100°.

a Find the probability distribution for Y.

b Find the probability that the alarm will function when the temperature reaches 100°.

3.46 Construct probability histograms for the binomial probability distributions for $n = 5$, $p = .1$, .5, and .9. (Table 1, Appendix 3, will reduce the amount of calculation.) Notice the symmetry for $p = .5$ and the direction of skewness for $p = .1$ and .9.

3.47 Use Table 1, Appendix 3, to construct a probability histogram for the binomial probability distribution for $n = 20$ and $p = .5$. Notice that almost all the probability falls in the interval $5 \le y \le 15$.

3.48 A missile protection system consists of n radar sets operating independently, each with a probability of .9 of detecting a missile entering a zone that is covered by all of the units.

a If $n = 5$ and a missile enters the zone, what is the probability that exactly four sets detect the missile? At least one set?

b How large must n be if we require that the probability of detecting a missile that enters the zone be .999?

3.49 A manufacturer of floor wax has developed two new brands, A and B, which she wishes to subject to homeowners' evaluation to determine which of the two is superior. Both waxes, A and B, are applied to floor surfaces in each of 15 homes. Assume that there is actually no difference in the quality of the brands. What is the probability that ten or more homeowners would state a preference for

a brand A?

b either brand A or brand B?

3.50 In Exercise 2.151, you considered a model for the World Series. Two teams A and B play a series of games until one team wins four games. We assume that the games are played independently and that the probability that A wins any game is p. Compute the probability that the series lasts exactly five games. [*Hint:* Use what you know about the random variable, Y, the number of games that A wins among the first four games.]

3.51 In the 18th century, the Chevalier de Mere asked Blaise Pascal to compare the probabilities of two events. Below, you will compute the probability of the two events that, prior to contrary gambling experience, were thought by de Mere to be equally likely.

a What is the probability of obtaining at least one 6 in four rolls of a fair die?

b If a pair of fair dice is tossed 24 times, what is the probability of at least one double six?

3.52 The taste test for PTC (phenylthiocarbamide) is a favorite exercise in beginning human genetics classes. It has been established that a single gene determines whether or not an individual is a "taster." If 70% of Americans are "tasters" and 20 Americans are randomly selected, what is the probability that

a at least 17 are "tasters"?

b fewer than 15 are "tasters"?

3.53 Tay-Sachs disease is a genetic disorder that is usually fatal in young children. If both parents are carriers of the disease, the probability that their offspring will develop the disease is approximately .25. Suppose that a husband and wife are both carriers and that they have three children. If the outcomes of the three pregnancies are mutually independent, what are the probabilities of the following events?

a All three children develop Tay-Sachs.

b Only one child develops Tay-Sachs.

c The third child develops Tay-Sachs, given that the first two did not.

3.54 Suppose that Y is a binomial random variable based on n trials with success probability p and consider $Y^* = n - Y$.

a Argue that for $y^* = 0, 1, \ldots, n$

$$P(Y^* = y^*) = P(n - Y = y^*) = P(Y = n - y^*).$$

b Use the result from part (a) to show that

$$P(Y^* = y^*) = \binom{n}{n - y^*} p^{n-y^*} q^{y^*} = \binom{n}{y^*} q^{y^*} p^{n-y^*}.$$

c The result in part (b) implies that Y^* has a binomial distribution based on n trials and "success" probability $p^* = q = 1 - p$. Why is this result "obvious"?

3.55 Suppose that Y is a binomial random variable with $n > 2$ trials and success probability p. Use the technique presented in Theorem 3.7 and the fact that $E\{Y(Y-1)(Y-2)\} = E(Y^3) - 3E(Y^2) + 2E(Y)$ to derive $E(Y^3)$.

3.56 An oil exploration firm is formed with enough capital to finance ten explorations. The probability of a particular exploration being successful is .1. Assume the explorations are independent. Find the mean and variance of the number of successful explorations.

3.57 Refer to Exercise 3.56. Suppose the firm has a fixed cost of \$20,000 in preparing equipment prior to doing its first exploration. If each successful exploration costs \$30,000 and each unsuccessful exploration costs \$15,000, find the expected total cost to the firm for its ten explorations.

3.58 A particular sale involves four items randomly selected from a large lot that is known to contain 10% defectives. Let Y denote the number of defectives among the four sold. The purchaser of the items will return the defectives for repair, and the repair cost is given by $C = 3Y^2 + Y + 2$. Find the expected repair cost. [*Hint:* The result of Theorem 3.6 implies that, for any random variable Y, $E(Y^2) = \sigma^2 + \mu^2$.]

3.59 Ten motors are packaged for sale in a certain warehouse. The motors sell for \$100 each, but a double-your-money-back guarantee is in effect for any defectives the purchaser may receive. Find the expected net gain for the seller if the probability of any one motor being defective is .08. (Assume that the quality of any one motor is independent of that of the others.)

3.60 A particular concentration of a chemical found in polluted water has been found to be lethal to 20% of the fish that are exposed to the concentration for 24 hours. Twenty fish are placed in a tank containing this concentration of chemical in water.

a Find the probability that exactly 14 survive.

b Find the probability that at least 10 survive.

c Find the probability that at most 16 survive.

d Find the mean and variance of the number that survive.

3.61 Of the volunteers donating blood in a clinic, 80% have the Rhesus (Rh) factor present in their blood.

a If five volunteers are randomly selected, what is the probability that at least one does not have the Rh factor?

b If five volunteers are randomly selected, what is the probability that at most four have the Rh factor?

c What is the smallest number of volunteers who must be selected if we want to be at least 90% certain that we obtain at least five donors with the Rh factor?

3.62 Goranson and Hall (1980) explain that the probability of detecting a crack in an airplane wing is the product of p_1, the probability of inspecting a plane with a wing crack; p_2, the probability of inspecting the detail in which the crack is located; and p_3, the probability of detecting the damage.

a What assumptions justify the multiplication of these probabilities?

b Suppose $p_1 = .9$, $p_2 = .8$, and $p_3 = .5$ for a certain fleet of planes. If three planes are inspected from this fleet, find the probability that a wing crack will be detected on at least one of them.

***3.63** Consider the binomial distribution with n trials and $P(S) = p$.

a Show that $\dfrac{p(y)}{p(y-1)} = \dfrac{(n-y+1)p}{yq}$ for $y = 1, 2, \ldots, n$. Equivalently, for $y = 1, 2, \ldots, n$, the equation $p(y) = \dfrac{(n-y+1)p}{yq}p(y-1)$ gives a recursive relationship between the probabilities associated with successive values of Y.

b If $n = 90$ and $p = .04$, use the above relationship to find $P(Y < 3)$.

c Show that $\dfrac{p(y)}{p(y-1)} = \dfrac{(n-y+1)p}{yq} > 1$ if $y < (n+1)p$, that $\dfrac{p(y)}{p(y-1)} < 1$ if $y > (n+1)p$, and that $\dfrac{p(y)}{p(y-1)} = 1$ if $(n+1)p$ is an integer and $y = (n+1)p$. This establishes that $p(y) > p(y-1)$ if y is small ($y < (n+1)p$) and $p(y) < p(y-1)$ if y is large ($y > (n+1)p$). Thus, successive binomial probabilities increase for a while and decrease from then on.

d Show that the value of y assigned the largest probability is equal to the greatest integer less than or equal to $(n+1)p$. If $(n+1)p = m$ for some integer m, then $p(m) = p(m-1)$.

***3.64** Consider an extension of the situation discussed in Example 3.10. If there are n trials in a binomial experiment and we observe y_0 "successes," show that $P(Y = y_0)$ is maximized when $p = y_0/n$. Again, we are determining (in general this time) the value of p that maximizes the probability of the value of Y that we actually observed.

***3.65** Refer to Exercise 3.64. The *maximum likelihood estimator* for p is Y/n (note that Y is the binomial random variable, not a particular value of it).

a Derive $E(Y/n)$. In Chapter 9, we will see that this result implies that Y/n is an *unbiased* estimator for p.

b Derive $V(Y/n)$. What happens to $V(Y/n)$ as n gets large?

3.5 The Geometric Probability Distribution

The random variable with the geometric probability distribution is associated with an experiment that shares some of the characteristics of a binomial experiment. This experiment also involves identical and independent trials, each of which can result in one of two outcomes: success or failure. The probability of success is equal to p and is constant from trial to trial. However, instead of the number of successes that occur in n trials, the geometric random variable Y is the number of the trial on which the first success occurs. Thus, the experiment consists of a series of trials that concludes with the first success. Consequently, the experiment could end with the first trial if a success is observed on the very first trial, or the experiment could go on indefinitely.

The sample space S for the experiment contains the countably infinite set of sample points:

E_1:	S	(success on first trial)
E_2:	FS	(failure on first, success on second)
E_3:	FFS	(first success on the third trial)
E_4:	$FFFS$	(first success on the fourth trial)
$\vdots$		
E_k:	$\underbrace{FFFF\dots F}_{k-1}S$	(first success on the k^{th} trial)
$\vdots$		

Because the random variable Y is the number of trials up to and including the first success, the events $(Y = 1)$, $(Y = 2)$, and $(Y = 3)$ contain only the sample points E_1, E_2, and E_3, respectively. More generally, the numerical event $(Y = y)$ contains only E_y. Because the trials are independent, for any $y = 1, 2, 3, \dots$,

$$p(y) = P(Y = y) = P(E_y) = P(\underbrace{FFFF\dots F}_{y-1}S) = \underbrace{qqq\cdots q}_{y-1}\,p = q^{y-1}p.$$

DEFINITION 3.8

A random variable Y is said to have a *geometric probability distribution* if and only if

$$p(y) = q^{y-1}p, \qquad y = 1, 2, 3, \dots, \quad 0 \le p \le 1.$$

A probability histogram for $p(y)$, $p = .5$, is shown in Figure 3.5. Areas over intervals correspond to probabilities, as they did for the frequency distributions of data in Chapter 1, except that Y can assume only discrete values, $y = 1, 2, \dots, \infty$. That $p(y) \ge 0$ is obvious by inspection of the respective values. In Exercise 3.66 you will show that these probabilities add up to 1, as is required for any valid discrete probability distribution.

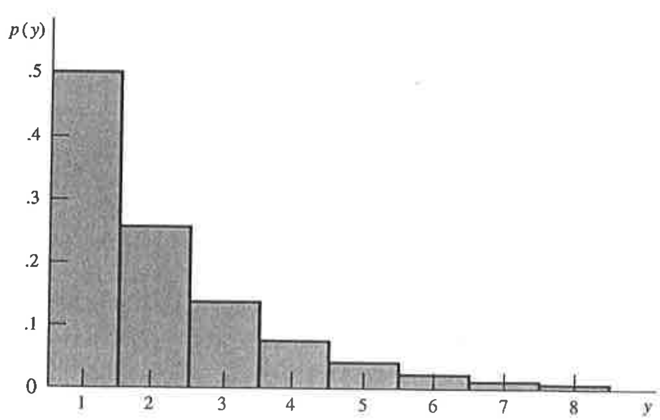

FIGURE 3.5 The geometric probability distribution, $p = .5$

The geometric probability distribution is often used to model distributions of lengths of waiting times. For example, suppose that a commercial aircraft engine is serviced periodically so that its various parts are replaced at different points in time and hence are of varying ages. Then the probability of engine malfunction, p, during any randomly observed one-hour interval of operation might be the same as for any other one-hour interval. The length of time prior to engine malfunction is the number of one-hour intervals, Y, until the first malfunction. (For this application, engine malfunction in a given one-hour period is defined as a success. Notice that, as in the case of the binomial experiment, either of the two outcomes of a trial can be defined as a success. Again, a "success" is not necessarily what would be considered to be "good" in everyday conversation.)

EXAMPLE **3.11** Suppose that the probability of engine malfunction during any one-hour period is $p = .02$. Find the probability that a given engine will survive two hours.

Solution Letting Y denote the number of one-hour intervals until the first malfunction, we have

$$P(\text{survive two hours}) = P(Y \geq 3) = \sum_{y=3}^{\infty} p(y).$$

Because $\sum_{y=1}^{\infty} p(y) = 1$,

$$\begin{aligned} P(\text{survive two hours}) &= 1 - \sum_{y=1}^{2} p(y) \\ &= 1 - p - qp = 1 - .02 - (.98)(.02) = .9604. \end{aligned}$$

■

If you examine the formula for the geometric distribution given in Definition 3.8, you will see that larger values of p (and hence smaller values of q) lead to higher probabilities for the smaller values of Y and hence lower probabilities for the larger values of Y. Thus, the mean value of Y appears to be inversely proportional to p. As we show in the next theorem, the mean of a random variable with a geometric distribution is actually equal to $1/p$.

THEOREM **3.8** If Y is a random variable with a geometric distribution,

$$\mu = E(Y) = \frac{1}{p} \quad \text{and} \quad \sigma^2 = V(Y) = \frac{1-p}{p^2}.$$

Proof

$$E(Y) = \sum_{y=1}^{\infty} yq^{y-1}p = p\sum_{y=1}^{\infty} yq^{y-1}.$$

This series might seem to be difficult to sum directly. Actually, it can be summed easily if we take into account that, for $y \geq 1$,

$$\frac{d}{dq}(q^y) = yq^{y-1},$$

and, hence,

$$\frac{d}{dq}\left(\sum_{y=1}^{\infty} q^y\right) = \sum_{y=1}^{\infty} yq^{y-1}.$$

(The interchanging of derivative and sum here can be justified.) Substituting, we obtain

$$E(Y) = p\sum_{y=1}^{\infty} yq^{y-1} = p\frac{d}{dq}\left(\sum_{y=1}^{\infty} q^y\right).$$

The latter sum is the geometric series, $q + q^2 + q^3 + \cdots$, which is equal to $q/(1-q)$ (see Appendix A1.11). Therefore,

$$E(Y) = p\frac{d}{dq}\left(\frac{q}{1-q}\right) = p\left[\frac{1}{(1-q)^2}\right] = \frac{p}{p^2} = \frac{1}{p}.$$

To summarize, our approach is to express a series that cannot be summed directly as the derivative of a series for which the sum can be readily obtained. Once we evaluate the more easily handled series, we differentiate to complete the process.

The derivation of the variance is left as Exercise 3.85.

EXAMPLE 3.12 If the probability of engine malfunction during any one-hour period is $p = .02$ and Y denotes the number of one-hour intervals until the first malfunction, find the mean and standard deviation of Y.

Solution As in Example 3.11, it follows that Y has a geometric distribution with $p = .02$. Thus, $E(Y) = 1/p = 1/(.02) = 50$, and we expect to wait quite a few hours before encountering a malfunction. Further, $V(Y) = .98/.0004 = 2450$, and it follows that the standard deviation of Y is $\sigma = \sqrt{2450} = 49.497$. ■

Although the computation of probabilities associated with geometric random variables can be accomplished by evaluating a single value or partial sums associated with a geometric series, these probabilities can also be found using various computer software packages. If Y has a geometric distribution with success probability p, $P(Y = y_0) = p(y_0)$ can be found by using the R (or S-Plus) command `dgeom(y0-1,p)`, whereas $P(Y \leq y_0)$ is found by using the R (or S-Plus) command `pgeom(y0-1,p)`. For example, the R (or S-Plus) command `pgeom(1,0.02)` yields the value for

$P(Y \leq 2)$ that was implicitly used in Example 3.11. Note that the argument in these commands is the value $y_0 - 1$, not the value y_0. This is because some authors prefer to define the geometric distribution to be that of the random variable $Y^\star =$ *the number of failures before the first success*. In our formulation, the geometric random variable Y is interpreted as *the number of the trial on which the first success occurs*. In Exercise 3.88, you will see that $Y^\star = Y - 1$. Due to this relationship between the two versions of geometric random variables, $P(Y = y_0) = P(Y - 1 = y_0 - 1) = P(Y^\star = y_0 - 1)$. R computes probabilities associated with $Y^\star$, explaining why the arguments for `dgeom` and `pgeom` are $y_0 - 1$ instead of y_0.

The next example, similar to Example 3.10, illustrates how knowledge of the geometric probability distribution can be used to estimate an unknown value of p, the probability of a success.

EXAMPLE 3.13 Suppose that we interview successive individuals working for the large company discussed in Example 3.10 and stop interviewing when we find the first person who likes the policy. If the fifth person interviewed is the first one who favors the new policy, find an estimate for p, the true but unknown proportion of employees who favor the new policy.

Solution If Y denotes the number of individuals interviewed until we find the first person who likes the new retirement plan, it is reasonable to conclude that Y has a geometric distribution for some value of p. Whatever the true value for p, we conclude that the probability of observing the first person in favor of the policy on the fifth trial is

$$P(Y = 5) = (1 - p)^4 p.$$

We will use as our estimate for p the value that maximizes the probability of observing the value that we *actually observed* (the first success on trial 5).

To find the value of p that maximizes $P(Y = 5)$, we again observe that the value of p that maximizes $P(Y = 5) = (1 - p)^4 p$ is the same as the value of p that maximizes $\ln[(1 - p)^4 p] = [4 \ln(1 - p) + \ln(p)]$.

If we take the derivative of $[4 \ln(1 - p) + \ln(p)]$ with respect to p, we obtain

$$\frac{d[4 \ln(1 - p) + \ln(p)]}{dp} = \frac{-4}{1 - p} + \frac{1}{p}.$$

Setting this derivative equal to 0 and solving, we obtain $p = 1/5$.

Because the second derivative of $[4 \ln(1 - p) + \ln(p)]$ is negative when $p = 1/5$, it follows that $[4 \ln(1 - p) + \ln(p)]$ [and $P(Y = 5)$] is *maximized* when $p = 1/5$. Our estimate for p, based on observing the first success on the fifth trial is $1/5$.

Perhaps this result is a little more surprising than the answer we obtained in Example 3.10 where we estimated p on the basis of observing 6 in favor of the new plan in a sample of size 20. Again, this is an example of the use of the *method of maximum likelihood* that will be studied in more detail in Chapter 9. ■

Problems on Geometric Random Variables

...ment in these
tho... efer to
the number of
...om variable Y
...rs. In Exercise
...wo versions of
...$^* = y_0 - 1$). R
...nts for dgeom

...wledge of the
...wn value of p,

...large company
...rst person who
...favors the new
...employees who

...irst person who
...has a geometric
...onclude that the
...fifth trial is

...lity of observing

...ve that the value
...value of p that

...p, we obtain

...e when $p = 1/5$,
...when $p = 1/5$.
...trial is $1/5$.
... we obtained in
...favor of the new
...of the *method of*
...9.

Exercises

Warning: R's dgeom is $f(x) = (1-p)^x p$ with $x = 0, 1, 2,$ rather than our $f(x) = (1-p)^{x-1} \cdot p$. R's geometric random variable counts the number of unsuccessful trials before success, so is one less than in Hogg-Tanis and WMS.

3.66 Suppose that Y is a random variable with a geometric distribution. Show that

a $\sum_y p(y) = \sum_{y=1}^{\infty} q^{y-1} p = 1$.

b $\dfrac{p(y)}{p(y-1)} = q$, for $y = 2, 3, \ldots$. This ratio is less than 1, implying that the geometric probabilities are monotonically decreasing as a function of y. If Y has a geometric distribution, what value of Y is the most likely (has the highest probability)?

3.67 Suppose that 30% of the applicants for a certain industrial job possess advanced training in computer programming. Applicants are interviewed sequentially and are selected at random from the pool. Find the probability that the first applicant with advanced training in programming is found on the fifth interview.

.07203 dgeom(4,.3)

3.68 Refer to Exercise 3.67. What is the expected number of applicants who need to be interviewed in order to find the first one with advanced training?

3.33

3.69 About six months into George W. Bush's second term as president, a Gallup poll indicated that a near record (low) level of 41% of adults expressed "a great deal" or "quite a lot" of confidence in the U.S. Supreme Court (http://www.gallup.com/poll/content/default.aspx?ci=17011, June 2005). Suppose that you conducted your own telephone survey at that time and randomly called people and asked them to describe their level of confidence in the Supreme Court. Find the probability distribution for Y, the number of calls until the first person is found who *does not* express "a great deal" or "quite a lot" of confidence in the U.S. Supreme Court.

3.70 An oil prospector will drill a succession of holes in a given area to find a productive well. The probability that he is successful on a given trial is .2.

a What is the probability that the third hole drilled is the first to yield a productive well?

b If the prospector can afford to drill at most ten wells, what is the probability that he will fail to find a productive well?

3.71 Let Y denote a geometric random variable with probability of success p.

a Show that for a positive integer a,

$$P(Y > a) = q^a.$$

b Show that for positive integers a and b,

$$P(Y > a + b | Y > a) = q^b = P(Y > b).$$

This result implies that, for example, $P(Y > 7|Y > 2) = P(Y > 5)$. Why do you think this property is called the *memoryless* property of the geometric distribution?

c In the development of the distribution of the geometric random variable, we assumed that the experiment consisted of conducting identical and independent trials until the first success was observed. In light of these assumptions, why is the result in part (b) "obvious"?

3.72 Given that we have already tossed a balanced coin ten times and obtained zero heads, what is the probability that we must toss it at least two more times to obtain the first head?

3.73 A certified public accountant (CPA) has found that nine of ten company audits contain substantial errors. If the CPA audits a series of company accounts, what is the probability that the first account containing substantial errors

a is the third one to be audited?

b will occur on or after the third audited account?

3.74 Refer to Exercise 3.73. What are the mean and standard deviation of the number of accounts that must be examined to find the first one with substantial errors?

3.75 The probability of a customer arrival at a grocery service counter in any one second is equal to .1. Assume that customers arrive in a random stream and hence that an arrival in any one second is independent of all others. Find the probability that the first arrival

a will occur during the third one-second interval.

b will not occur until at least the third one-second interval.

3.76 If Y has a geometric distribution with success probability .3, what is the largest value, y_0, such that $P(Y > y_0) \geq .1$?

3.77 If Y has a geometric distribution with success probability p, show that

$$P(Y = \text{ an odd integer }) = \frac{p}{1-q^2}.$$

3.78 Of a population of consumers, 60% are reputed to prefer a particular brand, A, of toothpaste. If a group of randomly selected consumers is interviewed, what is the probability that exactly five people have to be interviewed to encounter the first consumer who prefers brand A? At least five people?

3.79 In responding to a survey question on a sensitive topic (such as "Have you ever tried marijuana?"), many people prefer not to respond in the affirmative. Suppose that 80% of the population have not tried marijuana and all of those individuals will truthfully answer no to your question. The remaining 20% of the population have tried marijuana and 70% of those individuals will lie. Derive the probability distribution of Y, the number of people you would need to question in order to obtain a single affirmative response.

3.80 Two people took turns tossing a fair die until one of them tossed a 6. Person A tossed first, B second, A third, and so on. Given that person B threw the first 6, what is the probability that B obtained the first 6 on her second toss (that is, on the fourth toss overall)?

3.81 How many times would you expect to toss a balanced coin in order to obtain the first head?

3.82 Refer to Exercise 3.70. The prospector drills holes until he finds a productive well. How many holes would the prospector expect to drill? Interpret your answer intuitively.

3.83 The secretary in Exercises 2.121 and 3.16 was given n computer passwords and tries the passwords at random. Exactly one of the passwords permits access to a computer file. Suppose now that the secretary selects a password, tries it, and—if it does not work—puts it back in with the other passwords before randomly selecting the next password to try (not a very clever secretary!). What is the probability that the correct password is found on the sixth try?

3.84 Refer to Exercise 3.83. Find the mean and the variance of Y, the number of the trial on which the correct password is first identified.

***3.85** Find $E[Y(Y-1)]$ for a geometric random variable Y by finding $d^2/dq^2\left(\sum_{y=1}^{\infty} q^y\right)$. Use this result to find the variance of Y.

***3.86** Consider an extension of the situation discussed in Example 3.13. If we observe y_0 as the value for a geometric random variable Y, show that $P(Y = y_0)$ is maximized when $p = 1/y_0$. Again, we are determining (in general this time) the value of p that maximizes the probability of the value of Y that we actually observed.

it will be rejected? What is the expected number of defectives in the sample of size 5? What is the variance of the number of defectives in the sample of size 5?

Solution Let Y equal the number of defectives in the sample. Then $N = 20$, $r = 4$, and $n = 5$. The lot will be rejected if $Y = 2, 3$, or 4. Then

$$\begin{aligned} P(\text{rejecting the lot}) &= P(Y \geq 2) = p(2) + p(3) + p(4) \\ &= 1 - p(0) - p(1) \\ &= 1 - \frac{\binom{4}{0}\binom{16}{5}}{\binom{20}{5}} - \frac{\binom{4}{1}\binom{16}{4}}{\binom{20}{5}} \\ &= 1 - .2817 - .4696 = .2487. \end{aligned}$$

The mean and variance of the number of defectives in the sample of size 5 are

$$\mu = \frac{(5)(4)}{20} = 1 \quad \text{and} \quad \sigma^2 = 5\left(\frac{4}{20}\right)\left(\frac{20-4}{20}\right)\left(\frac{20-5}{20-1}\right) = .632. \quad \blacksquare$$

Example 3.17 involves sampling a lot of N industrial products, of which r are defective. The random variable of interest is Y, the number of defectives in a sample of size n. As noted in the beginning of this section, Y possesses an approximately binomial distribution when N is large and n is relatively small. Consequently, we would expect the probabilities assigned to values of Y by the hypergeometric distribution to approach those assigned by the binomial distribution as N becomes large and r/N, the fraction defective in the population, is held constant and equal to p. You can verify this expectation by using limit theorems encountered in your calculus courses to show that

$$\lim_{N\to\infty} \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} = \binom{n}{y} p^y (1-p)^{n-y},$$

where

$$\frac{r}{N} = p.$$

(The proof of this result is omitted.) Hence, for a fixed fraction defective $p = r/N$, the hypergeometric probability function converges to the binomial probability function as N becomes large.

Problems on Hypergeometric Random Variables

Exercises

3.102 An urn contains ten marbles, of which five are green, two are blue, and three are red. Three marbles are to be drawn from the urn, one at a time without replacement. What is the probability that all three marbles drawn will be green?

3.103 A warehouse contains ten printing machines, four of which are defective. A company selects five of the machines at random, thinking all are in working condition. What is the probability that all five of the machines are nondefective?

amp… of size 5?
…e

= 4, and $n = 5$.

(4)

…e of size 5 are

) = .632. ■

s, of which r are
ctives in a sample
an approximately
Consequently, we
…rgeometric distri-
N becomes large
…nt and equal to p.
…d in your calculus

…ctive $p = r/N$, the
robability function

…d three are red. Three
What is the probability

…e. A company selects
What is the probability

3.104 Twenty identical looking packets of white power are such that 15 contain cocaine and 5 do not. Four packets were randomly selected, and the contents were tested and found to contain cocaine. Two additional packets were selected from the remainder and sold by undercover police officers to a single buyer. What is the probability that the 6 packets randomly selected are such that the first 4 all contain cocaine and the 2 sold to the buyer do not?

3.105 In southern California, a growing number of individuals pursuing teaching credentials are choosing paid internships over traditional student teaching programs. A group of eight candidates for three local teaching positions consisted of five who had enrolled in paid internships and three who enrolled in traditional student teaching programs. All eight candidates appear to be equally qualified, so three are randomly selected to fill the open positions. Let Y be the number of internship trained candidates who are hired.

a Does Y have a binomial or hypergeometric distribution? Why?

b Find the probability that two or more internship trained candidates are hired.

c What are the mean and standard deviation of Y?

3.106 Refer to Exercise 3.103. The company repairs the defective ones at a cost of \$50 each. Find the mean and variance of the total repair cost.

3.107 A group of six software packages available to solve a linear programming problem has been ranked from 1 to 6 (best to worst). An engineering firm, unaware of the rankings, randomly selected and then purchased two of the packages. Let Y denote the number of packages purchased by the firm that are ranked 3, 4, 5, or 6. Give the probability distribution for Y.

3.108 A shipment of 20 cameras includes 3 that are defective. What is the minimum number of cameras that must be selected if we require that $P(\text{at least 1 defective}) \geq .8$?

3.109 Seed are often treated with fungicides to protect them in poor draining, wet environments. A small-scale trial, involving five treated and five untreated seeds, was conducted prior to a large-scale experiment to explore how much fungicide to apply. The seeds were planted in wet soil, and the number of emerging plants were counted. If the solution was not effective and four plants actually sprouted, what is the probability that

a all four plants emerged from treated seeds?

b three or fewer emerged from treated seeds?

c at least one emerged from untreated seeds?

3.110 A corporation is sampling without replacement for $n = 3$ firms to determine the one from which to purchase certain supplies. The sample is to be selected from a pool of six firms, of which four are local and two are not local. Let Y denote the number of nonlocal firms among the three selected.

.6 **a** $P(Y = 1)$.

.8 **b** $P(Y \geq 1)$.

.8 **c** $P(Y \leq 1)$.

3.111 Specifications call for a thermistor to test out at between 9000 and 10,000 ohms at 25° Celcius. Ten thermistors are available, and three of these are to be selected for use. Let Y denote the number among the three that do not conform to specifications. Find the probability distributions for Y (in tabular form) under the following conditions:

a Two thermistors do not conform to specifications among the ten that are available.

b Four thermistors do not conform to specifications among the ten that are available.

a)

y	0	1	2
p(y)	14/30	14/30	2/30

b)

y	0	1	2	3
p(y)	5/30	15/30	9/30	1/30

3.112 Used photocopy machines are returned to the supplier, cleaned, and then sent back out on lease agreements. Major repairs are not made, however, and as a result, some customers receive malfunctioning machines. Among eight used photocopiers available today, three are malfunctioning. A customer wants to lease four machines immediately. To meet the customer's deadline, four of the eight machines are randomly selected and, without further checking, shipped to the customer. What is the probability that the customer receives

- **a** no malfunctioning machines?
- **b** at least one malfunctioning machine?

3.113 A jury of 6 persons was selected from a group of 20 potential jurors, of whom 8 were African American and 12 were white. The jury was supposedly randomly selected, but it contained only 1 African American member. Do you have any reason to doubt the randomness of the selection?

3.114 Refer to Exercise 3.113. If the selection process were really random, what would be the mean and variance of the number of African American members selected for the jury?

3.115 Suppose that a radio contains six transistors, two of which are defective. Three transistors are selected at random, removed from the radio, and inspected. Let Y equal the number of defectives observed, where $Y = 0, 1,$ or 2. Find the probability distribution for Y. Express your results graphically as a probability histogram.

y	0	1	2
$f(y)$	1/5	3/5	1/5

3.116 Simulate the experiment described in Exercise 3.115 by marking six marbles or coins so that two represent defectives and four represent nondefectives. Place the marbles in a hat, mix, draw three, and record Y, the number of defectives observed. Replace the marbles and repeat the process until $n = 100$ observations of Y have been recorded. Construct a relative frequency histogram for this sample and compare it with the population probability distribution (Exercise 3.115).

3.117 In an assembly-line production of industrial robots, gearbox assemblies can be installed in one minute each if holes have been properly drilled in the boxes and in ten minutes if the holes must be redrilled. Twenty gearboxes are in stock, 2 with improperly drilled holes. Five gearboxes must be selected from the 20 that are available for installation in the next five robots.

$Y :=$ number of improperly drilled gearboxes

- **a** Find the probability that all 5 gearboxes will fit properly. $P(Y=0)=.553$
- **b** Find the mean, variance, and standard deviation of the time it takes to install these 5 gearboxes.

$T=9Y+5$, $E(T)=9.5$ $V(T)=28.755$ $\sigma=5.362$

3.118 Five cards are dealt at random and without replacement from a standard deck of 52 cards. What is the probability that the hand contains all 4 aces if it is known that it contains at least 3 aces?

3.119 Cards are dealt at random and without replacement from a standard 52 card deck. What is the probability that the second king is dealt on the fifth card?

***3.120** The sizes of animal populations are often estimated by using a capture–tag–recapture method. In this method k animals are captured, tagged, and then released into the population. Some time later n animals are captured, and Y, the number of tagged animals among the n, is noted. The probabilities associated with Y are a function of N, the number of animals in the population, so the observed value of Y contains information on this unknown N. Suppose that $k = 4$ animals are tagged and then released. A sample of $n = 3$ animals is then selected at random from the same population. Find $P(Y = 1)$ as a function of N. What value of N will *maximize* $P(Y = 1)$?

Exercises

Problems on Poisson Random Variables

3.121 Let Y denote a random variable that has a Poisson distribution with mean $\lambda = 2$. Find

a $P(Y = 4)$.
b $P(Y \geq 4)$.
c $P(Y < 4)$.
d $P(Y \geq 4 | Y \geq 2)$.

3.122 Customers arrive at a checkout counter in a department store according to a Poisson distribution at an average of seven per hour. During a given hour, what are the probabilities that

.818 **a** no more than three customers arrive?
.9927 **b** at least two customers arrive?
.1277 **c** exactly five customers arrive?

3.123 The random variable Y has a Poisson distribution and is such that $p(0) = p(1)$. What is $p(2)$?

3.124 Approximately 4% of silicon wafers produced by a manufacturer have fewer than two large flaws. If Y, the number of flaws per wafer, has a Poisson distribution, what proportion of the wafers have more than five large flaws? [*Hint:* Use Table 3, Appendix 3.]

$T = 10Y$
$E(T) = 70$
$V(T) = 700$

3.125 Refer to Exercise 3.122. If it takes approximately ten minutes to serve each customer, find the mean and variance of the total service time for customers arriving during a 1-hour period. (Assume that a sufficient number of servers are available so that no customer must wait for service.) Is it likely that the total service time will exceed 2.5 hours?

$P(T > 150) = .002$

3.126 Refer to Exercise 3.122. Assume that arrivals occur according to a Poisson process with an average of seven per hour. What is the probability that exactly two customers arrive in the two-hour period of time between

$\frac{14^2 e^{-14}}{2!}$ for both

a 2:00 P.M. and 4:00 P.M. (one continuous two-hour period)?
b 1:00 P.M. and 2:00 P.M. or between 3:00 P.M. and 4:00 P.M. (two separate one-hour periods that total two hours)?

3.127 The number of typing errors made by a typist has a Poisson distribution with an average of four errors per page. If more than four errors appear on a given page, the typist must retype the whole page. What is the probability that a randomly selected page does not need to be retyped?

3.128 Cars arrive at a toll both according to a Poisson process with mean 80 cars per hour. If the attendant makes a one-minute phone call, what is the probability that at least 1 car arrives during the call?

***3.129** Refer to Exercise 3.128. How long can the attendant's phone call last if the probability is at least .4 that no cars arrive during the call?

3.130 A parking lot has two entrances. Cars arrive at entrance I according to a Poisson distribution at an average of three per hour and at entrance II according to a Poisson distribution at an average of four per hour. What is the probability that a total of three cars will arrive at the parking lot in a given hour? (Assume that the numbers of cars arriving at the two entrances are independent.)

3.131 The number of knots in a particular type of wood has a Poisson distribution with an average of 1.5 knots in 10 cubic feet of the wood. Find the probability that a 10-cubic-foot block of the wood has at most 1 knot.

3.132 The mean number of automobiles entering a mountain tunnel per two-minute period is one. An excessive number of cars entering the tunnel during a brief period of time produces a hazardous situation. Find the probability that the number of autos entering the tunnel during a two-minute period exceeds three. Does the Poisson model seem reasonable for this problem?

3.133 Assume that the tunnel in Exercise 3.132 is observed during ten two-minute intervals, thus giving ten independent observations $Y_1, Y_2, \ldots, Y_{10}$, on the Poisson random variable. Find the probability that $Y > 3$ during at least one of the ten two-minute intervals.

Make the pmfs and see they are similar.

3.134 Consider a binomial experiment for $n = 20$, $p = .05$. Use ~~Table 1~~ R, Appendix 3, to calculate the binomial probabilities for $Y = 0, 1, 2, 3$, and 4. Calculate the same probabilities by using the Poisson approximation with $\lambda = np$. Compare.

3.135 A salesperson has found that the probability of a sale on a single contact is approximately .03. If the salesperson contacts 100 prospects, what is the approximate probability of making at least one sale?

3.136 Increased research and discussion have focused on the number of illnesses involving the organism *Escherichia coli* (10257:H7), which causes a breakdown of red blood cells and intestinal hemorrhages in its victims (http://www.hsus.org/ace/11831, March 24, 2004). Sporadic outbreaks of *E.coli* have occurred in Colorado at a rate of approximately 2.4 per 100,000 for a period of two years.

a If this rate has not changed and if 100,000 cases from Colorado are reviewed for this year, what is the probability that at least 5 cases of *E.coli* will be observed?

b If 100,000 cases from Colorado are reviewed for this year and the number of *E.coli* cases exceeded 5, would you suspect that the state's mean *E.coli* rate has changed? Explain.

3.137 The probability that a mouse inoculated with a serum will contract a certain disease is .2. Using the Poisson approximation, find the probability that at most 3 of 30 inoculated mice will contract the disease.

3.138 Let Y have a Poisson distribution with mean λ. Find $E[Y(Y-1)]$ and then use this to show that $V(Y) = \lambda$.

3.139 In the daily production of a certain kind of rope, the number of defects per foot Y is assumed to have a Poisson distribution with mean $\lambda = 2$. The profit per foot when the rope is sold is given by X, where $X = 50 - 2Y - Y^2$. Find the expected profit per foot.

***3.140** A store owner has overstocked a certain item and decides to use the following promotion to decrease the supply. The item has a marked price of \$100. For each customer purchasing the item during a particular day, the owner will reduce the price by a factor of one-half. Thus, the first customer will pay \$50 for the item, the second will pay \$25, and so on. Suppose that the number of customers who purchase the item during the day has a Poisson distribution with mean 2. Find the expected cost of the item at the end of the day. [*Hint:* The cost at the end of the day is $100(1/2)^Y$, where Y is the number of customers who have purchased the item.]

3.141 A food manufacturer uses an extruder (a machine that produces bite-size cookies and snack food) that yields revenue for the firm at a rate of \$200 per hour when in operation. However, the extruder breaks down an average of two times every day it operates. If Y denotes the number of breakdowns per day, the daily revenue generated by the machine is $R = 1600 - 50Y^2$. Find the expected daily revenue for the extruder.

***3.142** Let $p(y)$ denote the probability function associated with a Poisson random variable with mean λ.

a Show that the ratio of successive probabilities satisfies $\dfrac{p(y)}{p(y-1)} = \dfrac{\lambda}{y}$, for $y = 1, 2, \ldots$.

b For which values of y is $p(y) > p(y-1)$?