7.3 Geometric Distributions

In some board games, you cannot move forward until you roll a specific number, which could take several tries. Manufacturers of products such as switches, relays, and hard drives need to know how many operations their products can perform before failing. In some sports competitions, the winner is the player who scores the most points before missing a shot. In each of these situations, the critical quantity is the **waiting time** or **waiting period**—the number of trials before a specific outcome occurs.

INVESTIGATE & INQUIRE: Simulating Waiting Times

To get out of jail in the game of MONOPOLY®, you must either roll doubles or pay the bank $50. Design a simulation to find the probabilities of getting out of jail in x rolls of the two dice.

1. Select a random-number generator to simulate the selection process.
2. Decide how to simplify the selection process. Decide, also, whether the full situation needs to be simulated or whether a proportion of the trials would be sufficient.
3. Design each trial so that it simulates the actual situation. Determine whether your simulation tool must be reset or replaced after each trial to properly correspond to rolling two dice.
4. Set up a method to record the frequency of each outcome. Record the number of failures before a success finally occurs.
5. Combine your results with those of your classmates, if necessary.
6. Use these results to calculate an empirical probability for each outcome and the expected waiting time before a success.
7. Reflect on the results. Do they accurately represent the expected number of failures before success?
8. Compare your simulation and its results with those of other students in your class. Which simulation do you think worked best? Explain the reasons for your choice.

The simulation above models a **geometric distribution**. Like binomial distributions, trials in a geometric distribution have only two possible outcomes, success or failure, whose probabilities do not change from one trial to the next. However, the random variable for a geometric distribution is the *waiting time*, the number of unsuccessful independent trials before success occurs. Having different random variables causes significant differences between binomial and geometric distributions.

Example 1 Getting out of Jail in MONOPOLY®

a) Calculate the probability distribution for getting out of jail in MONOPOLY® in x rolls of the dice.

b) Estimate the expected number of rolls before getting out of jail.

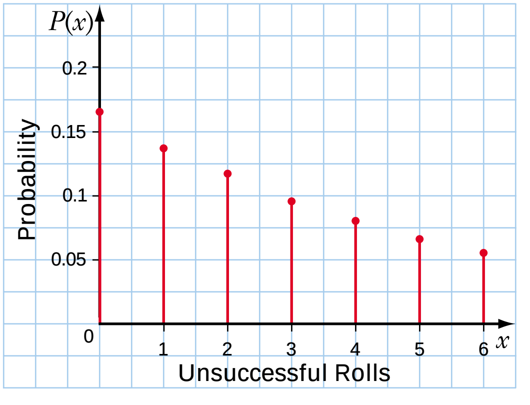
Solution 1 Using Pencil and Paper

a) The random variable is the number of unsuccessful rolls before you get out of jail. You can get out of jail by rolling doubles, and $P(\text{doubles}) = \frac{6}{36}$. So, for each independent roll,

$$p = \frac{6}{36} = \frac{1}{6} \quad \text{and} \quad q = 1 - \frac{1}{6} = \frac{5}{6}$$

You can apply the product rule to find the probability of successive independent events (see section 6.3). Thus, each unsuccessful roll preceding the successful one adds a factor of $\frac{5}{6}$ to the probability.

Unsuccessful Rolls (Waiting Time), x	Probability, $P(x)$
0	$\frac{1}{6} = 0.166\,66\ \ldots$
1	$\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = 0.138\,88\ \ldots$
2	$\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right) = 0.115\,74\ \ldots$
3	$\left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right) = 0.096\,45\ \ldots$
4	$\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right) = 0.080\,37\ \ldots$
…	…



This distribution theoretically continues forever since one possible outcome is that the player never rolls doubles. However, the probability for a waiting time decreases markedly as the waiting time increases. Although this distribution is an infinite geometric series, its terms still sum to 1 since they represent the probabilities of all possible outcomes.

b) Calculate the first six terms. If these terms approach zero rapidly, the sum of these terms will give a rough first approximation for the expectation.

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} xP(x) \\
&= (0)(0.16666\ldots) + (1)(0.13888\ldots) + (2)(0.11574\ldots) + 3(0.09645\ldots) \\
&\quad + (4)(0.08037\ldots) + (5)(0.06698\ldots) + \ldots \\
&> 0 + 0.13888 \;\; + 0.23148 \;\; + 0.28935 + 0.32150 + 0.33489 + \ldots \\
&> 1.3
\end{aligned}
$$

Clearly, the six terms are not approaching zero rapidly. All you can conclude is that the expected number of rolls before getting out of jail in MONOPOLY® is definitely more than 1.3.

Solution 2 Using a Graphing Calculator

a) You can use the calculator's lists to display the probabilities. Clear lists L1 to L4. For a start, enter the integers from 0 to 39 into L1. The **seq(function** in the LIST OPS menu provides a convenient way to enter these numbers.

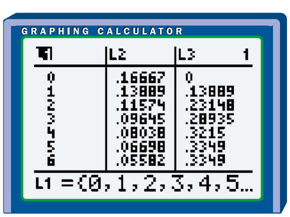
seq(A,A,0,39)→L1

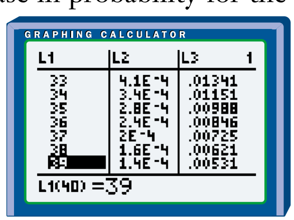
Next, use the geometric probability density function in the DISTR menu to calculate the probability of each value of x. The **geometpdf(function** has the syntax

geometpdf(*probability of success*, *number of trial on which first success occurs*)

Since x is the number of trials *before* success occurs, the number of the trial on which the first success occurs is $x + 1$. Therefore, enter geometpdf(1/6,L1+1)→L2.

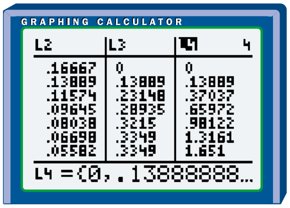
Notice the dramatic decrease in probability for the higher values of x.

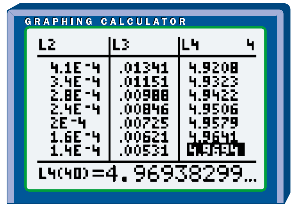




b) In L3, calculate the value of $xP(x)$ with the simple formula L1 × L2. You can sum these values to get a reasonable estimate for the expectation. To determine the accuracy of this estimate, it is helpful to look at the cumulative or running total of the $xP(x)$ values in L3.

To do this, in L4, select 6:cumSum(from the LIST OPS menu and type L3).





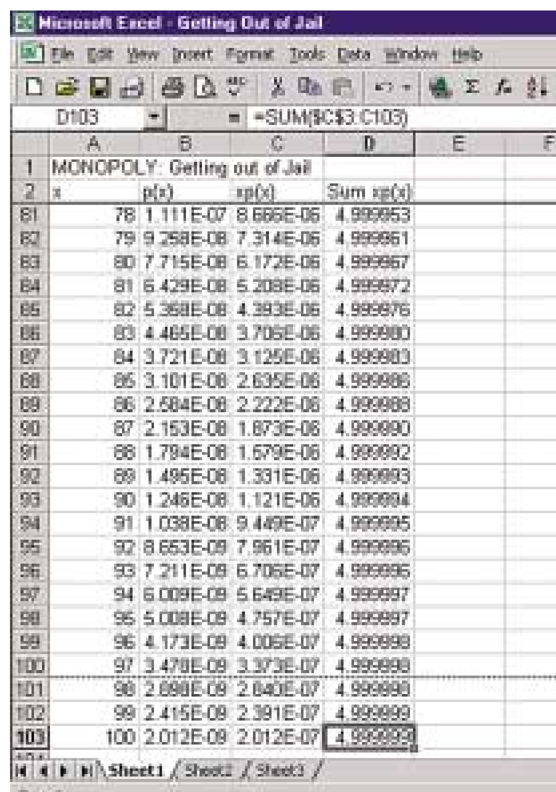
Note how the running total of $xP(x)$ in L4 increases more and more slowly toward the end of the list, suggesting that the infinite series for the expectation will total to a little more than 4.97. Thus, 5.0 would be a reasonable estimate for the number of trials before rolling doubles. You can check this estimate by performing similar calculations for waiting times of 100 trials or more.

Solution 3 Using a Spreadsheet

a) Open a new spreadsheet. Enter the headings x, p(x), and xp(x) in columns A, B, and C. Use the **Fill feature** to enter a sequence of values of the random variable x in column A, starting with 0 and going up to 100. Calculate the probability $P(x)$ for each value of the random variable x by entering the formula (5/6)^A3*(1/6) in cell B3 and then copying it down the rest of the column.

b) You can calculate $xP(x)$ in column C by entering the formula A3*B3 in cell C3 and then copying it down the rest of the column. Next, calculate the cumulative expected values in column D using the **SUM function** (with absolute cell references for the first cell) in cell D3 and copying it down the column.

Note that for $x = 50$ the cumulative expected value is over 4.99. By $x = 100$, it has reached 4.999 999 and is increasing extremely slowly. Thus, 5.000 00 is an accurate estimate for the expected number of trials before rolling doubles.



	A	B	C	D	E	F
1	MONOPOLY: Getting out of Jail					
2	x	p(x)	xp(x)	Sum xp(x)		
81	78	1.111E-07	8.666E-06	4.999853		
82	79	9.258E-08	7.314E-06	4.999861		
83	80	7.715E-08	6.172E-06	4.999867		
84	81	6.429E-08	5.208E-06	4.999872		
85	82	5.358E-08	4.393E-06	4.999876		
86	83	4.465E-08	3.706E-06	4.999880		
87	84	3.721E-08	3.125E-06	4.999883		
88	85	3.101E-08	2.635E-06	4.999886		
89	86	2.584E-08	2.222E-06	4.999888		
90	87	2.153E-08	1.873E-06	4.999890		
91	88	1.794E-08	1.579E-06	4.999892		
92	89	1.495E-08	1.331E-06	4.999893		
93	90	1.246E-08	1.121E-06	4.999894		
94	91	1.038E-08	9.449E-07	4.999895		
95	92	8.653E-09	7.961E-07	4.999896		
96	93	7.211E-09	6.706E-07	4.999896		
97	94	6.009E-09	5.649E-07	4.999897		
98	95	5.008E-09	4.757E-07	4.999897		
99	96	4.173E-09	4.006E-07	4.999898		
100	97	3.478E-09	3.373E-07	4.999898		
101	98	2.898E-09	2.840E-07	4.999898		
102	99	2.415E-09	2.391E-07	4.999899		
103	100	2.012E-09	2.012E-07	4.999899		

You can use the method in Example 1 to show that the probability of success after a waiting time of x failures is

Probability in a Geometric Distribution

$P(x) = q^x p$,
where p is the probability of success in each single trial and q is the probability of failure.

The expectation of a geometric distribution is the sum of an infinite series. Using calculus, it is possible to show that this expectation converges to a simple formula.

Expectation for a Geometric Distribution

$$E(X) = \sum_{x=0}^{\infty} xP(x) = \frac{q}{p}$$

Project Prep

Techniques for calculating expected values will be useful for your probability distributions project.

Example 2 Expectation of Geometric Distribution

Use the formula for the expectation of a geometric distribution to evaluate the accuracy of the estimates in Example 1.

Solution

$$E(X) = \frac{q}{p} = \frac{\frac{5}{6}}{\frac{1}{6}} = 5$$

For this particular geometric distribution, the simple manual estimate in Example 1 is accurate only to an order of magnitude. However, the calculator and the spreadsheet estimates are much more accurate.

Example 3 Basketball Free Throws

Jamaal has a success rate of 68% for scoring on free throws in basketball. What is the expected waiting time before he misses the basket on a free throw?

Solution

Here, the random variable is the number of trials before Jamaal misses on a free throw. For calculating the waiting time, a success is Jamaal *failing* to score. Thus,

$$q = 0.68 \quad \text{and} \quad p = 1 - 0.68 = 0.32$$

Using the expectation formula for the geometric distribution,

$$E(X) = \frac{q}{p} = \frac{0.68}{0.32} = 2.1$$

The expectation is that Jamaal will score on 2.1 free throws before missing.

Example 4 Traffic Management

Suppose that an intersection you pass on your way to school has a traffic light that is green for 40 s and then amber or red for a total of 60 s.

a) What is the probability that the light will be green when you reach the intersection at least once a week?

b) What is the expected number of days before the light is green when you reach the intersection?

Solution

a) Each trial is independent with

$$p = \frac{40}{100} = 0.40 \quad \text{and} \quad q = 0.60$$

There are five school days in a week. To get a green light on one of those five days, your waiting time must be four days or less.

$$P(0 \le x \le 4) = 0.40 + (0.60)(0.40) + (0.60)^2(0.40) + (0.60)^3(0.40) + (0.60)^4(0.40) = 0.92$$

The probability of the light being green when you reach the intersection at least once a week is 0.92.

b) $E(X) = \frac{q}{p}$

$= \frac{0.60}{0.40}$

$= 1.5$

The expected waiting time before catching a green light is 1.5 days.

Key Concepts

- A geometric distribution has a specified number of independent trials with two possible outcomes, success or failure. The random variable is the number of unsuccessful outcomes before a success occurs.
- The probability of success after a waiting time of x failures is $P(x) = q^x p$, where p is the probability of success in each single trial and q is the probability of failure.
- The expectation of a geometric distribution is $E(X) = \frac{q}{p}$.
- To simulate a geometric experiment, you must ensure that the probability on a single trial is accurate for the situation and that each trial is independent. Summarize the results by calculating probabilities and the expected waiting time.

Communicate Your Understanding

1. Describe how the graph in Example 1 differs from those of the uniform and binomial distributions.
2. Consider this question: What is the expected number of failures in 100 launches of a rocket that has a failure rate of 1.5%? Explain why this problem does not fit a geometric distribution and how it could be rewritten so that it does.

Practise

A

1. Which of the following situations is modelled by a geometric distribution? Explain your reasoning.
   - **a)** rolling a die until a 6 shows
   - **b)** counting the number of hearts when 13 cards are dealt from a deck
   - **c)** predicting the waiting time when standing in line at a bank
   - **d)** calculating the probability of a prize being won within the first 3 tries
   - **e)** predicting the number of successful launches of satellites this year

Data in Action

In 2000, there were 82 successful launches of satellites and 3 failures. Launching a satellite usually costs between $75 million and $600 million.

2. Prepare a table and a graph for six trials of a geometric distribution with
   a) $p = 0.2$
   b) $p = 0.5$

Apply, Solve, Communicate

B

3. For a 12-sided die,
   a) what is the probability that the first 10 will be on the third roll?
   b) what is the expected waiting time until a 1 is rolled?

4. The odds in favour of a Pythag-Air-US Airlines flight being on time are 3:1.
   a) What is the probability that this airline's next eight flights will be on time?
   b) What is the expected waiting time before a flight delay?

5. **Communication** To finish a board game, Sarah needed to land on the last square by rolling a sum of 2 with two dice. She was dismayed that it took her eight tries. Should she have been surprised? Explain.

6. In a TV game show, the grand prize is randomly hidden behind one of three doors. On each show, the finalist gets to choose one of the doors. What is the probability that no finalists will win a grand prize on four consecutive shows?

7. **Application** A teacher provides pizza for his class if they earn an A-average on any test. The probability of the class getting an A-average on one of his tests is 8%.
   a) What is the probability that the class will earn a pizza on the fifth test?
   b) What is the probability that the class will not earn a pizza for the first seven tests?
   c) What is the expected waiting time before the class gets a pizza?

8. Minh has a summer job selling replacement windows by telephone. Of the people he calls, nine out of ten hang up before he can give a sales pitch.
   a) What is the probability that, on a given day, Minh's first sales pitch is on his 12th call?
   b) What is the expected number of hang-ups before Minh can do a sales pitch?

9. Despite its name, Zippy Pizza delivers only 40% of its pizzas on time.
   a) What is the probability that its first four deliveries will be late on any given day?
   b) What is the expected number of pizza deliveries before one is on time?

10. A poll indicated that 34% of the population agreed with a recent policy paper issued by the government.
    a) What is the probability that the pollster would have to interview five people before finding a supporter of the policy?
    b) What is the expected waiting time before the pollster interviews someone who agrees with the policy?

11. Suppose that 1 out of 50 cards in a scratch-and-win promotion gives a prize.
    a) What is the probability of winning on your fourth try?
    b) What is the probability of winning within your first four tries?
    c) What is the expected number of cards you would have to try before winning?

12. A top NHL hockey player scores on 93% of his shots in a shooting competition.
    a) What is the probability that the player will not miss the goal until his 20th try?
    b) What is the expected number of shots before he misses?

13. **Application** A computer manufacturer finds that 1.5% of its chips fail quality-control testing.

a) What is the probability that one of the first five chips off the line will be defective?

b) What is the expected waiting time for a defective chip?

14. **Inquiry/Problem Solving** Three friends of an avid golfer have each given her a package of 5 balls for her birthday. The three packages are different brands. The golfer keeps all 15 balls in her golf bag and picks one at random at the start of each round.

a) Design a simulation to determine the waiting time before the golfer has tried all three brands. Assume that the golfer does not lose any of the balls.

b) Use the methods described in this section to calculate the expected waiting time before the golfer has tried the three different brands.

15. The Big K cereal company has randomly placed one of a set of seven different collector cards in each box of its Krakked Korn cereal. Each card is equally likely.

a) Design a simulation to estimate how many boxes of Krakked Korn you would have to buy to get a complete set of cards.

b) Use the methods described in this section to calculate the average number of boxes of Krakked Korn you would have to buy to get a complete set of cards.

C

16. **Inquiry/Problem Solving** Consider a geometric distribution where the random variable is the number of the trial with the first success instead of the number of failures before a success. Develop formulas for the probabilities and expectation for this distribution.

17. **Communication** A manufacturer of computer parts lists a mean time before failure (MTBF) of 5.4 years for one of its hard drives. Explain why this specification is different from the expected waiting time of a geometric distribution. Could you use the MTBF to calculate the probability of the drive failing in any one-year period?

18. In a sequence of Bernoulli trials, what is the probability that the second success occurs on the fifth trial?

19. **Communication** The rack behind a coat-check counter collapses and 20 coats slip off their numbered hangers. When the first person comes to retrieve one of these coats, the clerk brings them out and holds them up one at a time for the customer to identify.

a) What is the probability that the clerk will find the customer's coat

i) on the first try?

ii) on the second try?

iii) on the third try?

iv) in fewer than 10 tries?

b) What is the expected number of coats the clerk will have to bring out before finding the customer's coat?

c) Explain why you cannot use a geometric distribution to calculate this waiting time.