

4.4: Geometric Distribution

There are four main characteristics of a geometric experiment.

1. A trial is repeated until a success occurs. Think of this as one or more Bernoulli trials with all failures except the last one, which is a success. In other words, you keep repeating what you are doing until the first success. Then you stop. For example, you throw a dart at a bullseye until you hit the bullseye. The first time you hit the bullseye is a "success," so you stop throwing the dart. It might take six tries until you hit the bullseye. You can think of the trials as failure, failure, failure, failure, failure, success, STOP. In theory, the number of trials could go on forever.
2. The repeated trials are independent of each other.
3. The probability, p , of a success and the probability, q , of a failure is the same for each trial. $p + q = 1$ and $q = 1 - p$. For example, the probability of rolling a three when you throw one fair die is $1/6$. This is true no matter how many times you roll the die. Suppose you want to know the probability of getting the first three on the fifth roll. On rolls one through four, you do not get a face with a three. The probability for each of the rolls is $q = 5/6$, the probability of a failure. The probability of getting a first three on the fifth roll is $(5/6)(5/6)(5/6)(5/6)(1/6) = 0.0804$
4. The random variable X represents the number of the trial in which the first success occurs. That is, X = the number of independent trials until the first success.

The following are additional attributes of the geometric distribution:

5. The random variable is discrete. The random variable may be defined in two ways depending upon the analyst's interest. In the example above for throwing a die, the question was "What is the probability that the first success will be on the fifth throw?" Alternatively, the question could be asked as "What is the probability that it takes four failures before a success?" We will see that each way of asking the question will alter the form of the geometric probability density function slightly and will change the mean and standard deviation of the geometric pdf.
6. Implicit in the random variable is that the probability of a success is constant and therefore so is the probability of a failure. For flipping a coin this is obvious, but in experiments that require skill, such as hitting a baseball or throwing a dart, one might consider that learning during the experiment would alter the probability of a success. The geometric distribution cannot capture "learning" thus the historical probability of success is assumed to be constant.
7. The geometric distribution is "memoryless." There are very few probability density functions that are what is known as "memoryless," and the Geometric distribution is the only one with a discrete random variable that is memoryless. As an example, historically Major League Baseball player Jones has a record of hitting the ball for at least an advance to first base with a probability of 0.20. Jones has not had a hit in his last 10 times at bat. What is the probability that Jones will get a hit in his third time at bat? The answer ignores his 10 previous failures. All events prior to the events in current time are irrelevant and thus are considered "memoryless."

Formally: $P(x = n + k | x \geq k + 1) = P(x = n)$, where k = number of previous failures

Jones's probability of a hit begins anew each time he comes to bat. This feature of the geometric distribution results in a curious result: Drawing parts from a manufacturing process to test for parts that are defective, the geometric distribution begins with a clean slate each time the tests begin with no consideration of previous test results. More on this when we get to the exponential probability density function.

? Exercise 4.4.1

You play a game of chance that you can either win or lose (there are no other possibilities) **until** you lose. Your probability of losing is $p = 0.57$. What is the probability that it takes five games until you lose? Let X = the number of games you play until you lose (includes the losing game). Then X takes on the values 1, 2, 3, ... (could go on indefinitely). The probability question is $P(x = 5)$.

Try It 4.4.1

You throw darts at a board until you hit the center area. Your probability of hitting the center area is $p = 0.17$. You want to find the probability that it takes eight throws until you hit the center. What values does X take on?

? Exercise 4.4.2

A safety engineer feels that 35% of all industrial accidents in the plant are caused by failure of employees to follow instructions. They decide to look at the accident reports (selected randomly and replaced in the pile after reading) **until** they find one that shows an accident caused by failure of employees to follow instructions. On average, how many reports would the safety engineer **expect** to look at until they find a report showing an accident caused by employee failure to follow instructions? What is the probability that the safety engineer will have to examine at least three reports until they find a report showing an accident caused by employee failure to follow instructions?

Let X = the number of accidents the safety engineer must examine **until** they find a report showing an accident caused by employee failure to follow instructions. X takes on the values 1, 2, 3, The first question asks you to find the **expected value** or the mean. The second question asks you to find $P(x \geq 3)$. ("At least" translates to a "greater than or equal to" symbol).

Answer

a. On average, how many reports would the safety engineer expect to review until they find a report showing an accident caused by something other than failure to follow instructions? This question asks you to find the expected value or the mean. This is the average number of failures from something other than following instructions. The formula for the mean of this geometric distribution is:

$$\mu = E(x) = \frac{1}{p} \approx 2.8 \quad (4.4.1)$$

Note that the mean does not need to be a whole number although the random variable is discrete and must be a counting number

b. What is the probability that the safety engineer will have to examine at least three reports until they find a report showing an accident caused by employee failure to follow instructions? This question is answered by first defining the random variable. Let X = the number of accidents the safety engineer must examine until they find a report showing an accident caused by employee failure to follow instructions. X can take on the values 1, 2, 3, 4, ... ∞ Unlike the binomial distribution with a fixed number of trials, the geometric distribution may have an infinite number of failed trials before a success occurs.

This second question asks you to find $P(x \geq 3)$. ("At least" translates to a "greater than or equal to" symbol.)

In a binomial distribution, we saw the solution to questions of "more than" or "less than" would be to calculate each probability individually and add from zero to three. If the probability of interest is "greater than or equal to" (\geq) the individual probabilities are added from zero to three and then subtracted from one.

$$P(x \geq 3) = 1 - P(x < 3) \quad (4.4.2)$$

Because we cannot add the full range of the geometric distribution random variable because it goes through infinity, an alternative solution is developed. An alternative geometric distribution pair of formulas provides solutions for questions asking for probabilities "more than" and "less than."

If the question is:

What is the probability it takes MORE THAN n events to get first success?

$$P(x > n) = (1 - p)^n \quad (4.4.3)$$

What is the probability it takes LESS THAN n event for success?

$$P(x < n) = 1 - (1 - p)^n \quad (4.4.4)$$

Solution

- a. On average, how many reports would the safety engineer **expect** to review until they find a report showing an accident caused by something other than failure to follow instructions? This question asks you to find the **expected value** or the mean. This is the average number of failures from something other than following instructions. The formula for the mean of this geometric distribution is:

$$\mu = E(x) = \frac{1}{p} \approx 2.8$$

Note that the mean does not need to be a whole number although the random variable is discrete and must be a counting number

- b. What is the probability that the safety engineer will have to examine at least three reports until they find a report showing an accident caused by employee failure to follow instructions? This question is answered by first defining the random variable. Let X = the number of accidents the safety engineer must examine **until** they find a report showing an accident caused by employee failure to follow instructions. X can take on the values 1, 2, 3, 4, ... ∞ . Unlike the binomial distribution with a fixed number of trials, the geometric distribution may have an infinite number of failed trials before a success occurs.

This second question asks you to find $P(x \geq 3)$ ("At least" translates to a "greater than or equal to" symbol.)

$$P(x \geq 3)$$

In a binomial distribution, we saw the solution to questions of "more than" or "less than" would be to calculate each probability individually and add from zero to three. If the probability of interest is "greater than or equal to" (\geq) the individual probabilities are added from zero to three and then subtracted from one.

$$P(x \geq 3) = 1 - P(x < 3)$$

$$P(x \geq 3) = 1 - P(x < 3)$$

Because we cannot add the full range of the geometric distribution random variable because it goes through infinity, an alternative solution is developed. An alternative geometric distribution pair of formulas provides solutions for questions asking for probabilities "more than" and "less than."

If the question is:

What is the probability it takes MORE THAN n events to get first success?

$$P(x > n) = (1 - p)^n$$

$$P(x > n) = (1 - p)^n$$

What is the probability it takes LESS THAN n event for success?

$$P(x < n) = 1 - (1 - p)^n$$

$$P(x < n) = 1 - (1 - p)^n$$

Try It 4.4.2

An instructor feels that 15% of students get below a C on their final exam. They decide to look at final exams (selected randomly and replaced in the pile after reading) until they find one that shows a grade below a C. We want to know the probability that the instructor will have to examine at least ten exams until they find one with a grade below a C. What is the probability question stated mathematically?

Exercise 4.4.3

An instructor feels that 15% of students get below a C on their final exam. They decide to look at final exams (selected randomly and replaced in the pile after reading) until they find one that shows a grade below a C. We want to know the probability that the instructor will have to examine at least ten exams until they find one with a grade below a C. What is the probability question stated mathematically?

An instructor feels that 15% of students get below a C on their final exam. They decide to look at final exams (selected randomly and replaced in the pile

- Let X = the number of _____ you must ask _____ one says yes.
- What values does X take on?
- What are p and q ?
- The probability question is $P(\text{_____})$.

Answer

- Let X = the number of **students** you must ask **until** one says yes.
- 1, 2, 3, ..., (total number of students)
- $p = 0.15$; $q = 0.85$
- $P(X \geq 10)$

Solution

- Let X = the number of **students** you must ask **until** one says yes.
- 1, 2, 3, ..., (total number of students)
- $p = 0.15$; $q = 0.85$
- $P(X \geq 10)$

Try It 4.4.1

You need to find a store that carries a special printer ink. You know that of the stores that carry printer ink, 10% of them carry the special ink. You randomly call each store until one has the ink you need. What are p and q ?

Notation for the Geometric: $G = \text{Geometric Probability Distribution Function}$

Read this as " X is a random variable with a **geometric distribution**." The parameter is p ; p = the probability of a success for each trial.

CASE I: Random Variable X Is Event of First Success

In this case we ask, "What is the probability that we will have some number x of events of interest to us of failures before a success?"

The geometric pdf tells us the probability that the first occurrence of success requires x number of failure independent trials, each with probability $(1-p)$. If the probability of success on each trial is p , then the probability that the x th trial (out of x trials) is the first success is:

$$P(X = x) = (1 - p)^{x-1} p \quad (4.4.5)$$

for $x = 1, 2, 3, \dots$

Like the binomial distribution, the geometric distribution has the parameters of the mean and standard deviation. The expected value of X , the mean for Case I, is $\mu = \frac{1}{p}$. This tells us how many failed trials to expect until we get the first success. This count includes in the count of trials the trial that results in success. The above form of the geometric distribution is used for modeling the number of trials until the first success. The number of trials includes the one that is a success: x = all trials including the one that is a success. This can be seen in the form of the formula. If X = number of trials including the success, then we must multiply the probability of failure, $(1 - p)$, times the number of failures, that is $x - 1$. The standard deviation of Case I of the geometric distribution is:

$$\sigma = \sqrt{\left(\frac{1}{p}\right) \left(\frac{1}{p} - 1\right)} \quad (4.4.6)$$

CASE II: Random Variable X Is Number of Failures BEFORE a Success

By contrast to Case I, the following form of the geometric distribution used for modeling number of failures until the first success is:

$$P(X = x) = (1 - p)^x p \quad (4.4.7)$$

for $x = 0, 1, 2, 3, \dots$

In this case the trial that is the success is not counted as a trial in the formula: x = number of failures. The expected value, the mean, of this distribution is $\mu = \frac{1-p}{p}$. This tells us how many failures to expect before we have a success. In either case, the sequence of probabilities is a geometric sequence. In Case II the standard deviation parameter is:

$$\sigma = \sqrt{\left(\frac{1-p}{p}\right)} \quad (4.4.8)$$

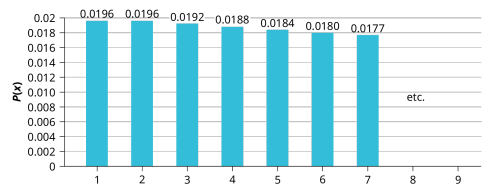


Figure 4.4.1: Geometric Distribution: $P = 0.02$ Common Ratio $r = 0.98$

The y -axis in Figure 4.4.1 contains the probability of x , and the x -axis is the random number components tested. For example, at $x = 1$ the probability it will be found to be defective is 0.0196. With two components tested, the probability the second component is defective is graphed at a probability of 0.0196 at $x = 2$ on the x axis. For the probability that the third component is defective we find $P(X = 3) = 0.019208$. (The first two are the same because of rounding in the computations.)

Notice in Figure 4.3 that the probabilities decline by the same step down with each change in the value of x . This increment is called the common ratio. This exists for the geometric probability distribution uniquely. The common ratio, called r , can be calculated by dividing any value by the previous value, e.g., $\frac{P(x=5)}{P(x=4)} = \frac{0.018447}{0.018823} = 0.98$

For this set of data, therefore, the common ratio is 0.98. The common ratio then multiplied by any other probability value will provide the next probability value in the sequence. For example, the probability that the sixth component tested is a failure is 0.018078. Check this using the formula from Case I. Now we have the $P(x = 6)$. Knowing this and the common ratio we can calculate the $P(x = 7)$ by simple multiplication:

$P(x = 7) = (0.018078) \times 0.98 = 0.017716$ the same value we found earlier by using the geometric probability distribution. This common ratio increment is the same ratio between every number and is called a geometric progression and thus the name for this probability density function. Once the common ratio is calculated, any $P(x = x_a)$ one desires to know can be easily found.

The number of components that you would expect to test until you find the first defective component is the mean, $\mu = 50$ for this case of defective components. The formula for the mean of the geometric distribution for the random variable defined as number of failures before first success is $\mu = \frac{1}{p} = \frac{1}{0.02} = 50$.

See Example 4.4.4 for an example where the geometric random variable is defined as number of trials until first success. The expected value of this formula for the geometric distribution will be different from this version of the distribution. Case II also has a variance but is changed from the Case I formula. This formula for the variance is:

$$P(X = x) = (1 - p)^{x-1} p \quad (4.4.9)$$

for $x = 1, 2, 3, \dots$

The formula for the variance is

$$\sigma^2 = \left(\frac{1}{p}\right) \left(\frac{1}{p} - 1\right) = \left(\frac{1}{0.02}\right) \left(\frac{1}{0.02} - 1\right) = 2,450 \quad (4.4.10)$$

The standard deviation is

$$\sigma = \sqrt{\left(\frac{1}{p}\right) \left(\frac{1}{p} - 1\right)} = \sqrt{\left(\frac{1}{0.02}\right) \left(\frac{1}{0.02} - 1\right)} = 49.5 \quad (4.4.11)$$

This tells us how many failures to expect before we have a success. In either case, the sequence of probabilities is a geometric sequence.

? Exercise 4.4.4

Assume that the probability of a defective computer component is 0.02. Components are randomly selected. Find the probability that the first defect is caused by the seventh component tested. How many components do you expect to test until one is found to be defective?

Let X = the number of computer components tested until the first defect is found.

X takes on the values 1, 2, 3, ... where $p = 0.02$.

Find $P(x = 7)$. Answer: $P(x = 7) = (1 - 0.02)^{7-1} \times 0.02 = 0.0177$.

The probability that the seventh component is the first defect is 0.0177.

The graph of $X \sim G(0.02)$ is:

The y -axis contains the probability of x , where X = the number of computer components tested. Notice that the probabilities decline by a common increment. This increment is the same ratio between each number and is called a geometric progression and thus the name for this probability density function.

See Example 4.4.5 for an example where the geometric random variable is defined as number of trials until first success. The expected value of this formula for the geometric will be different from this version of the distribution.

The formula for the variance is $\sigma^2 = \left(\frac{1}{p}\right) \left(\frac{1}{p} - 1\right) = \left(\frac{1}{0.02}\right) \left(\frac{1}{0.02} - 1\right) = 2,450$

The standard deviation is $\sigma = \sqrt{\left(\frac{1}{p}\right) \left(\frac{1}{p} - 1\right)} = \sqrt{\left(\frac{1}{0.02}\right) \left(\frac{1}{0.02} - 1\right)} = 49.5$

Try It 4.4.4

The probability of a defective steel rod is 0.01. Steel rods are selected at random. Find the probability that the first defect occurs on the ninth steel rod. Use the TI-83+ or TI-84 calculator to find the answer.

? Exercise 4.4.5

The lifetime risk of developing cancer is about one in 67 (1.5%). Let X = the number of people you ask before one says they have cancer. The random variable X in this case includes only the number of trials that were failures and does not count the trial that was a success in finding a person who had the disease. The appropriate formula for this random variable is the second one presented above. Then X is a discrete random variable with a geometric distribution:

- What is the probability that you ask 9 people before one says they have cancer? This is asking, what is the probability that you ask 9 people who say no and the tenth person says yes?
- What is the probability that you must ask 20 people?
- Find the (i) mean and (ii) standard deviation of X .

Answer

$$a. P(x=9) = (1-0.015)^9 \times 0.015 = 0.0131$$

$$b. P(x=20) = (1-0.015)^{19} \times 0.015 = 0.0113$$

$$c. \quad i. \text{ Mean } = \mu = \frac{(1-p)}{p} = \frac{(1-0.015)}{0.015} = 65.67$$

$$ii. \text{ Standard Deviation } = \sigma = \sqrt{\frac{(1-p)}{p^2}} = \sqrt{\frac{(1-0.015)}{0.015^2}} = 66.16$$

Solution

$$a. P(x=10) = \text{geometricpdf}(0.015, 10) = 0.0131 \quad P(x=10) = \text{geometricpdf}(0.015, 10) = 0.0131$$

$$P(x=10) = \text{geometricpdf}(0.015, 10) = 0.0131$$

$$b. P(x=20) = \text{geometricpdf}(0.015, 20) = 0.0113 \quad P(x=20) = \text{geometricpdf}(0.015, 20) = 0.0113$$

$$P(x=20) = \text{geometricpdf}(0.015, 20) = 0.0113$$

$$c. \text{ Mean } = \mu = \frac{1-p}{p} = \frac{1-0.015}{0.015} = 65.67$$

$$\text{Mean } = \mu = \frac{1}{p} = \frac{1}{0.015} = 66.67$$

$$\text{Standard Deviation } = \sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.015}{0.015^2}} = 66.16$$

$$\text{Standard Deviation } = \sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.015}{0.015^2}} = 66.16$$

Try It 4.4.5

The literacy rate for a nation measures the proportion of people age 15 and over who can read and write. The literacy rate for women in The United Colonies of Independence is 12%. Let X = the number of women you ask until one says that she is literate.

- What is the probability distribution of X ?
- What is the probability that you ask five women before one says she is literate?
- What is the probability that you must ask ten women?

? Exercise 4.4.6

A baseball player has a batting average of 0.320. This is the general probability that he gets a hit each time he is at bat.

- What is the probability that he gets his first hit in the third trip to bat?
- In this case the sequence is failure, failure success. How many trips to bat do you expect the hitter to need before getting a hit?

Answer

$$a. P(x=3) = (1-0.32)^{3-1} \times 0.32 = 0.1480$$

$$b. \mu = \frac{1}{p} = \frac{1}{0.320} = 3.125 \approx 3$$

This is simply the expected value of successes and therefore the mean of the distribution.

Try It 4.4.6

A runner has a probability of 0.426 to come in first in a race.

- What is the probability that they will come in first in the fourth race?
- How many races do you expect the runner to need before they come in first?

? Exercise 4.4.1

There is an 80% chance that a Dalmatian dog has 13 black spots. You go to a dog show and count the spots on Dalmatians. What is the probability that you will review the spots on 3 dogs before you find one that has 13 black spots?

Answer

$$P(x=3) = (1-0.80)^3 \times 0.80 = 0.0064$$

Try It 4.4.1

There is a 75% chance that a parrot is green. You go to a pet shop and observe the parrots. What is the probability that you will observe 4 parrots before you find one that is a green parrot?