



# Geometric Distribution

By:  
OpenStaxCollege

There are three main characteristics of a geometric experiment.

1. There are 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.
2. In theory, the number of trials could go on forever. There must be at least one trial.
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  $\frac{1}{6}$ . 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  $q = \frac{5}{6}$ , the probability of a failure.

The probability of getting a three on the fifth roll is  $\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = 0.0804$

$X$  = the number of independent trials until the first success.

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

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?

1, 2, 3, 4, ...  $n$ . It can go on indefinitely.

A safety engineer feels that 35% of all industrial accidents in her plant are caused by failure of employees to follow instructions. She decides to look at the accident reports (selected randomly and replaced in the pile after reading) **until** she finds 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 she finds 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 she finds a report showing an accident caused by employee failure to follow instructions?

Let  $X$  = the number of accidents the safety engineer must examine **until** she finds 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).

Try It

An instructor feels that 15% of students get below a C on their final exam. She decides to look at final exams (selected randomly and replaced in the pile after reading) until she finds 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 she finds one with a grade below a C. What is the probability question stated mathematically?

$$P(x \geq 10)$$

Suppose that you are looking for a student at your college who lives within five miles of you. You know that 55% of the 25,000 students do live within five miles of you. You randomly contact students from the college **until** one says he or she lives within five miles of you. What is the probability that you need to contact four people?

This is a geometric problem because you may have a number of failures before you have the one success you desire. Also, the probability of a success stays the same each time you ask a student if he or she lives within five miles of you. There is no definite number of trials (number of times you ask a student).

a. Let  $X$  = the number of \_\_\_\_\_ you must ask \_\_\_\_\_ one says yes.

a. Let  $X$  = the number of **students** you must ask **until** one says yes.

b. What values does  $X$  take on?

b. 1, 2, 3, ..., (total number of students)

c. What are  $p$  and  $q$ ?

## Geometric Distribution

c.  $p = 0.55$ ;  $q = 0.45$

d. The probability question is  $P(\text{_____})$ .

d.  $P(x = 4)$

Try It

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  $q$ ?

$$p = 0.1$$

$$q = 0.9$$

### **Notation for the Geometric: G = Geometric Probability Distribution Function**

$$X \sim G(p)$$

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.

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  $p = 0.02$ .  $X \sim G(0.02)$

Find  $P(x = 7)$ .  $P(x = 7) = 0.0177$ .

To find the probability that  $x = 7$ ,

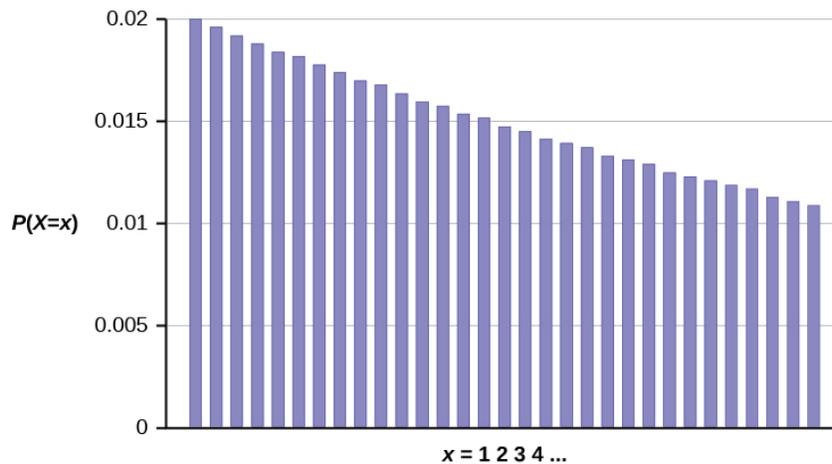
- Enter 2<sup>nd</sup>, DISTR
- Scroll down and select geometpdf(
- Press ENTER
- Enter 0.02, 7); press ENTER to see the result:  $P(x = 7) = 0.0177$

## Geometric Distribution

To find the probability that  $x \leq 7$ , follow the same instructions EXCEPT select E:geometcdf(as the distribution function.

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.

The number of components that you would expect to test until you find the first defective one is the mean,  $\mu = 50$ .

The formula for the mean is  $\mu = \frac{1}{p} = \frac{1}{0.02} = 50$

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

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.

$$P(x = 9) = 0.0092$$

The lifetime risk of developing pancreatic cancer is about one in 78 (1.28%). Let  $X$  = the number of people you ask until one says he or she has pancreatic cancer. Then  $X$  is a discrete random variable with a geometric distribution:  $X \sim G\left(\frac{1}{78}\right)$  or  $X \sim G(0.0128)$ .

1. What is the probability of that you ask ten people before one says he or she has pancreatic cancer?
2. What is the probability that you must ask 20 people?
3. Find the (i) mean and (ii) standard deviation of  $X$ .

1.  $P(x = 10) = \text{geometpdf}(0.0128, 10) = 0.0114$

2.  $P(x = 20) = \text{geometpdf}(0.0128, 20) = 0.01$

3.
  1. Mean  $= \mu = \frac{1}{p} = \frac{1}{0.0128} = 78$

2. Standard Deviation  $= \sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.0128}{0.0128^2}} \approx 77.6234$

Try It

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 Afghanistan is 12%. Let  $X$  = the number of Afghani women you ask until one says that she is literate.

1. What is the probability distribution of  $X$ ?
2. What is the probability that you ask five women before one says she is literate?
3. What is the probability that you must ask ten women?
4. Find the (i) mean and (ii) standard deviation of  $X$ .

1.  $X \sim G(0.12)$

2.  $P(x = 5) = \text{geometpdf}(0.12, 5) = 0.0720$

3.  $P(x = 10) = \text{geometpdf}(0.12, 10) = 0.0380$

4.
  1. Mean  $= \mu = \frac{1}{p} = \frac{1}{0.12} \approx 8.3333$

2. Standard Deviation  $= \sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.12}{0.12^2}} \approx 7.8174$

## References

“Millennials: A Portrait of Generation Next,” PewResearchCenter. Available online at <http://www.pewsocialtrends.org/files/2010/10/millennials-confident-connected-open-to-change.pdf> (accessed May 15, 2013).

“Millennials: Confident. Connected. Open to Change.” Executive Summary by PewResearch Social & Demographic Trends, 2013. Available online at <http://www.pewsocialtrends.org/2010/02/24/millennials-confident-connected-open-to-change/> (accessed May 15, 2013).

“Prevalence of HIV, total (% of populations ages 15-49),” The World Bank, 2013. Available online at <http://data.worldbank.org/indicator/>

SH.DYN.AIDS.ZS?order=wbapi\_data\_value\_2011+wbapi\_data\_value+wbapi\_data\_value-last&sort=desc (accessed May 15, 2013).

Pryor, John H., Linda DeAngelo, Laura Palucki Blake, Sylvia Hurtado, Serge Tran. *The American Freshman: National Norms Fall 2011*. Los Angeles: Cooperative Institutional Research Program at the Higher Education Research Institute at UCLA, 2011. Also available online at <http://heri.ucla.edu/PDFs/pubs/TFS/Norms/Monographs/TheAmericanFreshman2011.pdf> (accessed May 15, 2013).

“Summary of the National Risk and Vulnerability Assessment 2007/8: A profile of Afghanistan,” The European Union and ICON-Institute. Available online at [http://ec.europa.eu/europeaid/where/asia/documents/afgh\\_brochure\\_summary\\_en.pdf](http://ec.europa.eu/europeaid/where/asia/documents/afgh_brochure_summary_en.pdf) (accessed May 15, 2013).

“The World FactBook,” Central Intelligence Agency. Available online at <https://www.cia.gov/library/publications/the-world-factbook/geos/af.html> (accessed May 15, 2013).

“UNICEF reports on Female Literacy Centers in Afghanistan established to teach women and girls basic reading [sic] and writing skills,” UNICEF Television. Video available online at <http://www.unicefusa.org/assets/video/afghan-female-literacy-centers.html> (accessed May 15, 2013).

## Chapter Review

There are three characteristics of a geometric experiment:

1. There are one or more Bernoulli trials with all failures except the last one, which is a success.
2. In theory, the number of trials could go on forever. There must be at least one trial.
3. The probability,  $p$ , of a success and the probability,  $q$ , of a failure are the same for each trial.

In a geometric experiment, define the discrete random variable  $X$  as the number of independent trials until the first success. We say that  $X$  has a geometric distribution and write  $X \sim G(p)$  where  $p$  is the probability of success in a single trial.

The mean of the geometric distribution  $X \sim G(p)$  is  $\mu = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1}{p} \left( \frac{1}{p} - 1 \right)}$ .

### Formula Review

$X \sim G(p)$  means that the discrete random variable  $X$  has a geometric probability distribution with probability of success in a single trial  $p$ .

$X$  = the number of independent trials until the first success

$X$  takes on the values  $x = 1, 2, 3, \dots$

$p$  = the probability of a success for any trial

$q$  = the probability of a failure for any trial  $p + q = 1$

$q = 1 - p$

The mean is  $\mu = \frac{1}{p}$ .

The standard deviation is  $\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1}{p}(1-p)}$ .

*Use the following information to answer the next six exercises:* The Higher Education Research Institute at UCLA collected data from 203,967 incoming first-time, full-time freshmen from 270 four-year colleges and universities in the U.S. 71.3% of those students replied that, yes, they believe that same-sex couples should have the right to legal marital status. Suppose that you randomly select freshman from the study until you find one who replies “yes.” You are interested in the number of freshmen you must ask.

In words, define the random variable  $X$ .

$X$  = the number of freshmen selected from the study until one replied "yes" that same-sex couples should have the right to legal marital status.

$X \sim$  \_\_\_\_\_ (\_\_\_\_\_, \_\_\_\_\_)

What values does the random variable  $X$  take on?

1, 2, ...

Construct the probability distribution function (PDF). Stop at  $x = 6$ .

$x$	$P(x)$
1	
2	

$x$	$P(x)$
3	
4	
5	
6	

On average ( $\mu$ ), how many freshmen would you expect to have to ask until you found one who replies "yes?"

1.4

What is the probability that you will need to ask fewer than three freshmen?

### HOMEWORK

A consumer looking to buy a used red Miata car will call dealerships until she finds a dealership that carries the car. She estimates the probability that any independent dealership will have the car will be 28%. We are interested in the number of dealerships she must call.

1. In words, define the random variable  $X$ .
2. List the values that  $X$  may take on.
3. Give the distribution of  $X$ .  $X \sim \underline{\hspace{1cm}}(\underline{\hspace{1cm}}, \underline{\hspace{1cm}})$
4. On average, how many dealerships would we expect her to have to call until she finds one that has the car?
5. Find the probability that she must call at most four dealerships.
6. Find the probability that she must call three or four dealerships.

Suppose that the probability that an adult in America will watch the Super Bowl is 40%. Each person is considered independent. We are interested in the number of adults in America we must survey until we find one who will watch the Super Bowl.

1. In words, define the random variable  $X$ .
2. List the values that  $X$  may take on.
3. Give the distribution of  $X$ .  $X \sim \underline{\hspace{1cm}}(\underline{\hspace{1cm}}, \underline{\hspace{1cm}})$
4. How many adults in America do you expect to survey until you find one who will watch the Super Bowl?
5. Find the probability that you must ask seven people.
6. Find the probability that you must ask three or four people.

1.  $X$  = the number of adults in America who are surveyed until one says he or she will watch the Super Bowl.
2.  $X \sim G(0.40)$
3. 2.5
4. 0.0187
5. 0.2304

It has been estimated that only about 30% of California residents have adequate earthquake supplies. Suppose we are interested in the number of California residents we must survey until we find a resident who does **not** have adequate earthquake supplies.

1. In words, define the random variable  $X$ .
2. List the values that  $X$  may take on.
3. Give the distribution of  $X$ .  $X \sim \text{_____}(\text{_____,} \text{_____})$
4. What is the probability that we must survey just one or two residents until we find a California resident who does not have adequate earthquake supplies?
5. What is the probability that we must survey at least three California residents until we find a California resident who does not have adequate earthquake supplies?
6. How many California residents do you expect to need to survey until you find a California resident who **does not** have adequate earthquake supplies?
7. How many California residents do you expect to need to survey until you find a California resident who **does** have adequate earthquake supplies?

In one of its Spring catalogs, L.L. Bean® advertised footwear on 29 of its 192 catalog pages. Suppose we randomly survey 20 pages. We are interested in the number of pages that advertise footwear. Each page may be picked more than once.

1. In words, define the random variable  $X$ .
2. List the values that  $X$  may take on.
3. Give the distribution of  $X$ .  $X \sim \text{_____}(\text{_____,} \text{_____})$
4. How many pages do you expect to advertise footwear on them?
5. Is it probable that all twenty will advertise footwear on them? Why or why not?
6. What is the probability that fewer than ten will advertise footwear on them?
7. Reminder: A page may be picked more than once. We are interested in the number of pages that we must randomly survey until we find one that has footwear advertised on it. Define the random variable  $X$  and give its distribution.
8. What is the probability that you only need to survey at most three pages in order to find one that advertises footwear on it?
9. How many pages do you expect to need to survey in order to find one that advertises footwear?

1.  $X$  = the number of pages that advertise footwear
2.  $X$  takes on the values 0, 1, 2, ..., 20
3.  $X \sim B(20, \frac{29}{192})$
4. 3.02
5. No
6. 0.9997
7.  $X$  = the number of pages we must survey until we find one that advertises footwear.  $X \sim G(\frac{29}{192})$
8. 0.3881
9. 6.6207 pages

Suppose that you are performing the probability experiment of rolling one fair six-sided die. Let  $F$  be the event of rolling a four or a five. You are interested in how many times you need to roll the die in order to obtain the first four or five as the outcome.

- $p$  = probability of success (event  $F$  occurs)
  - $q$  = probability of failure (event  $F$  does not occur)
1. Write the description of the random variable  $X$ .
  2. What are the values that  $X$  can take on?
  3. Find the values of  $p$  and  $q$ .
  4. Find the probability that the first occurrence of event  $F$  (rolling a four or five) is on the second trial.

Ellen has music practice three days a week. She practices for all of the three days 85% of the time, two days 8% of the time, one day 4% of the time, and no days 3% of the time. One week is selected at random. What values does  $X$  take on?

0, 1, 2, and 3

The World Bank records the prevalence of HIV in countries around the world. According to their data, "Prevalence of HIV refers to the percentage of people ages 15 to 49 who are infected with HIV."

"Prevalence of HIV, total (% of populations ages 15-49)," The World Bank, 2013.

Available online at [http://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?order=wbapi\\_data\\_value\\_2011+wbapi\\_data\\_value+wbapi\\_data\\_value-last&sort=desc](http://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?order=wbapi_data_value_2011+wbapi_data_value+wbapi_data_value-last&sort=desc) (accessed May 15, 2013).

In South Africa, the prevalence of HIV is 17.3%. Let  $X$  = the number of people you test until you find a person infected with HIV.

1. Sketch a graph of the distribution of the discrete random variable  $X$ .

2. What is the probability that you must test 30 people to find one with HIV?
3. What is the probability that you must ask ten people?
4. Find the (i) mean and (ii) standard deviation of the distribution of  $X$ .

According to a recent Pew Research poll, 75% of millennials (people born between 1981 and 1995) have a profile on a social networking site. Let  $X$  = the number of millennials you ask until you find a person without a profile on a social networking site.

1. Describe the distribution of  $X$ .
2. Find the (i) mean and (ii) standard deviation of  $X$ .
3. What is the probability that you must ask ten people to find one person without a social networking site?
4. What is the probability that you must ask 20 people to find one person without a social networking site?
5. What is the probability that you must ask *at most* five people?

1.  $X \sim G(0.25)$

2. 1. Mean =  $\mu = \frac{1}{p} = \frac{1}{0.25} = 4$

2. 2. Standard Deviation =  $\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.25}{0.25^2}} \approx 3.4641$

3.  $P(x = 10) = \text{geompdf}(0.25, 10) = 0.0188$

4.  $P(x = 20) = \text{geompdf}(0.25, 20) = 0.0011$

5.  $P(x \leq 5) = \text{geomcdf}(0.25, 5) = 0.7627$