4.4: Geometric Distribution

The geometric probability density function builds upon what we have learned from the binomial distribution. In this case the experiment continues until either a success or a failure occurs rather than for a set number of trials. 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.
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 $\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)=0.0804$
4. X = the number of independent trials until the first success.

Example 4.4.5

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 $P(x=5)$.

Exercise 4.4.5

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?

Example 4.4.6

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).

Exercise 4.4.6

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?

Example 4.4.7

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 approximately 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.

Answer

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

b. What values does X take on?

Answer

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

c. What are p and q ?

Answer

c. $p = 0.55; q = 0.45$

d. The probability question is P (_______).

Answer

d. $P(x = 4)$

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.

The Geometric Pdf tells us the probability that the first occurrence of success requires x number of independent trials, each with success probability p. If the probability of success on each trial is p, then the probability that the xth trial (out of x trials) is the first success is:

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

for $x = 1, 2, 3, \ldots$.
The expected value of X, the mean of this distribution, is $1/p$. This tells us how many trials we have to expect until we get the first success including in the count 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 $X - 1$.

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

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

for $x = 0, 1, 2, 3, \ldots$.
In this case the trial that is a success is not counted as a trial in the formula: x = number of failures. The expected value, 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.

Example 4.4.8

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 $P(x = 7)$. 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:

Figure 4.4.2

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.

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

The formula for the mean for the random variable defined as number of failures until first success is $\mu = \frac{1}{p} = \frac{1}{0.02} = 50$

See Example 4.4.9 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 $\sigma = \sqrt{\left(\frac{1}{p}\right)\left(\frac{1}{p} - 1\right)} = \sqrt{\left(\frac{1}{0.02}\right)\left(\frac{1}{0.02} - 1\right)} = 49.5$

The lifetime risk of developing pancreatic cancer is about one in 78 (1.28%). Let X = the number of people you ask before one says he or she has pancreatic 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: X ~ G $\left(\frac{1}{78}\right)$ or X ~ G (0.0128).

a. What is the probability of that you ask 9 people before one says he or she has pancreatic cancer? This is asking, what is the probability that you ask 9 people unsuccessfully and the tenth person is a success?
b. What is the probability that you must ask 20 people?
c. Find the (i) mean and (ii) standard deviation of *X*.

Answer

a. $P(x = 9) = (1 - 0.0128)^9 \cdot 0.0128 = 0.0114$

b. $P(x = 20) = (1 - 0.0128)^{19} \cdot 0.0128 = 0.01$

i. Mean = $\mu = 78.00$

ii. Standard Deviation = $\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.0128}{0.0128^2}} \approx 77.62$

Exercise 4.4.9

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.

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

Example 4.4.10

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?

Answer

$$P(x=3)=(1-0.32)^{3-1} \times .32=0.1480$$

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

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

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

Example 4.4.11

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

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

Footnotes

1 ”Prevalence of HIV, total (% of populations ages 15-49),” The World Bank, 2013. Available online at http://data.worldbank.org/indicator/...last&sort=desc (accessed May 15, 2013).