Probability and Statistics (IT302) Class No. 22
6^{th} October 2020 Tuesday 10:30AM - 11:00AM

Negative Binomial and Geometric Distributions

Let us consider an experiment where the properties are the same as those listed for a binomial experiment, with the exception that the trials will be repeated until a fixed number of successes occur. Therefore, instead of the probability of x successes in n trials, where n is fixed, we are now interested in the probability that the k^{th} success occurs on the x^{th} trial. Experiments of this kind are called negative binomial experiments.

As an illustration, consider the use of a drug that is known to be effective in 60% of the cases where it is used. The drug will be considered a success if it is effective in bringing some degree of relief to the patient. We are interested in finding the probability that the fifth patient to experience relief is the seventh patient to receive the drug during a given week. Designating a success by *S* and a failure by *F*, a possible order of achieving the desired result is *SFSSSFS*, which occurs with probability

$$(0.6)(0.4)(0.6)(0.6)(0.6)(0.4)(0.6) = (0.6)^5(0.4)^2.$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall.

Hypergeometric Distribution Related Analytical Problems

We could list all possible orders by rearranging the *F*'s and *S*'s except for the last outcome, which must be the fifth success. The total number of possible orders is equal to the number of partitions of the first six trials into two groups with 2 failures assigned to the one group and 4 successes assigned to the other group.

This can be done in $\binom{6}{4} = 15$ mutually exclusive ways. Hence, if X represents the outcome on which the fifth success occurs, then

$$P(X = 7) = \binom{6}{4}(0.6)^5(0.4)^2 = 0.1866.$$

What Is the Negative Binomial Random Variable?

The number X of trials required to produce k successes in a negative binomial experiment is called a **Negative Binomial Random Variable**, and its probability distribution is called the **Negative Binomial Distribution**. Since its probabilities depend on the number of successes desired and the probability of a success on a given trial, we shall denote them by $b^*(x; k, p)$.

To obtain the general formula for $b^*(x; k, p)$, consider the probability of a success on the x^{th} trial preceded by $k - 1$ successes and $x - k$ failures in some specified order. Since the trials are independent, we can multiply all the probabilities corresponding to each desired outcome. Each success occurs with probability p and each failure with probability $q = 1 - p$.

Therefore, the probability for the specified order ending in success is

$$p^{k-1}q^{x-k}p = p^k q^{x-k}$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall.

What Is the Negative Binomial Random Variable? Contd.

The total number of sample points in the experiment ending in a success, after the occurrence of $k-1$ successes and $x-k$ failures in any order, is equal to the number of partitions of $x-1$ trials into two groups with $k-1$ successes corresponding to one group and $x-k$ failures corresponding to the other group. This number is specified

by the term $\binom{x-1}{k-1}$, each mutually exclusive and occurring with equal probability $p^k q^{x-k}$. We obtain the general formula by multiplying $p^k q^{x-k}$ by $\binom{x-1}{k-1}$.

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall.

Negative Binomial Distribution

If repeated independent trials can result in a success with **probability p** and **a failure with probability q = 1−p**, then the **probability distribution of the random variable X**, the number of the trial on which the k^{th} success occurs, is

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k}, \quad x = k, k+1, k+2, \ldots.$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall.

Example 5.14

In an NBA (National Basketball Association) championship series, the team that wins four games out of seven is the winner. Suppose that teams A and B face each other in the championship games and that team A has probability 0.55 of winning a game over team B.

a) What is the probability that team A will win the series in 6 games?
b) What is the probability that team A will win the series?
c) If teams A and B were facing each other in a regional playoff series, which is decided by winning three out of five games, what is the probability that team A would win the series?

Solution :

(a) $b^*(6;4,0.55) = \binom{5}{3}0.55^4(1-0.55)^{6-4} = 0.1853$

(b) $P(\text{team } A \text{ wins the championship series})$ is

$$\begin{aligned} &b^*(4;4,0.55) + b^*(5;4,0.55) + b^*(6;4,0.55) + b^*(7;4,0.55) \\ &= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083. \end{aligned}$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall.

Example 5.14 Contd.

(c) $P(\text{team } A \text{ wins the playoff})$ is

$$b^*(3;3,0.55) + b^*(4;3,0.55) + b^*(5;3,0.55)$$
$$= 0.1664 + 0.2246 + 0.2021 = 0.5931.$$

The negative binomial distribution derives its name from the fact that each term in the expansion of $p^k(1-q)^{-k}$ corresponds to the values of $b^*(x; k, p)$ for $x = k, k+1, k+2, \ldots$. If we consider the special case of the negative binomial distribution where k=1, we have a probability distribution for the number of trials required for a single success.

An example would be the tossing of a coin until a head occurs. We might be interested in the probability that the first head occurs on the fourth toss. The negative binomial distribution reduces to the form $b^*(x; 1, p) = pq^{x-1}$, $x = 1, 2, 3, \ldots$

Since the successive terms constitute a geometric progression, it is customary to refer to this special case as the geometric distribution and denote its values by g(x; p)

Geometric Distribution

If repeated independent trials can result in a success with **probability p** and a **failure with probability q = 1−p**, then the probability distribution of the random variable X, the number of the trial on which the first success occurs, is

$$g(x; p) = pq^{x-1}, \qquad x = 1, 2, 3, \ldots .$$

Example 5.15

For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

Solution :

Using the geometric distribution with $x = 5$ and $p = 0.01$,

$$g(5; 0.01) = (0.01)(0.99)^4 = 0.0096.$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall.

Example 5.16

At a "busy time," a telephone exchange is very near capacity, so callers have difficulty placing their calls. It may be of interest to know the number of attempts necessary in order to make a connection. Suppose that we let $p = 0.05$ be the probability of a connection during a busy time. We are interested in knowing the probability that 5 attempts are necessary for a successful call.

Solution :

Using the geometric distribution with $x = 5$ and $p = 0.05$ yields

$$P(X = x) = g(5; 0.05) = (0.05)(0.95)^4 = 0.041.$$

Quite often, in applications dealing with the geometric distribution, the mean and variance are important. For example, in **Example 5.16**, the expected number of calls necessary to make a connection is quite important. The following theorem states without proof the mean and variance of the geometric distribution.

Theorem 5.3

The mean and variance of a random variable following the geometric distribution are

$$\mu = \frac{1}{p} \text{ and } \sigma^2 = \frac{1-p}{p^2}.$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall.

Additional Materials

Negative Binomial Distribution

- The negative binomial distribution describes a sequence of trials, each of which can have two outcomes (success or failure).
- Continue the trials indefinitely until we get r successes.
- The prototypical example is flipping a coin until we get r heads.
- Unlike the binomial distribution, we don't know the number of trials in advance.
- The geometric distribution is the case r = 1

Formula for the Negative Binomial Distribution

Fixed parameters:

p := probability of success on each trial

q := probability of failure = 1 − p

r := number of successes desired

Random variable:

Y := number of trials (for r successes)

- Probability distribution:

$$\boxed{p(y) = \binom{y-1}{r-1} p^r q^{y-r}, r \leq y < \infty}$$

Warning

- These are different p's!
- p is the probability of success on any given trial.
- p(y) is the probability of y trials overall.

Source : https://www.educator.com/media/qn-dl/probability-murray/Prob-14.pdf

Example I

You draw cards from a deck (with replacement) until you get four aces. What is the chance that you will draw exactly 20 times?

Solution :

$$
\begin{aligned}
p &= \frac{1}{13} \\
p(20) &= \binom{y-1}{r-1} p^r q^{y-r} \\
&= \binom{19}{3}\left(\frac{1}{13}\right)^4\left(\frac{12}{13}\right)^{16} = \boxed{\binom{19}{3}\left(\frac{12^{16}}{13^{20}}\right)}
\end{aligned}
$$

Source : https://www.educator.com/media/qn-dl/probability-murray/Prob-14.pdf

Example II

Each year the Akron Aardvarks have a 10% chance of winning the trophy in chinchilla grooming. Their trophy case has space for five trophies. Let Y be the number of years until their case is full. Find the mean and standard deviation of Y .

Solution:

- **Mean:**

$$\boxed{\mu = E(Y) = \frac{r}{p}}$$

- **Variance:**

$$\boxed{\sigma^2 = V(Y) = \frac{rq}{p^2}}$$

- **Standard deviation:**

$$\boxed{\sigma = \sqrt{V(Y)} = \frac{\sqrt{rq}}{p}}$$

$$\begin{aligned}
p &= \frac{1}{10} \\
q &= \frac{9}{10} \\
r &= 5 \\
\mu &= \frac{r}{p} = \boxed{50 \text{ years}} \\
\sigma^2 &= \frac{rq}{p^2} = \frac{5\frac{9}{10}}{\frac{1}{10^2}} = 450 \\
\sigma &= \sqrt{450} = 15\sqrt{2} \approx \boxed{21.21 \text{ years}}
\end{aligned}$$

Source : https://www.educator.com/media/qn-dl/probability-murray/Prob-14.pdf

Example III

You roll a die until you get four sixes (not necessarily consecutive). What is the mean and standard deviation of the number of rolls you will make? This is the negative binomial distribution with p =1/6 ; r = 4.

Solution :

$$\begin{aligned}\mu &= \frac{r}{p} \\ &= \boxed{24 \text{ rolls}} \\ \sigma^2 &= \frac{rq}{p^2} \\ &= \frac{4 \cdot \frac{5}{6}}{\frac{1}{36}} \\ &= 120 \\ \sigma &= \sqrt{120} = \boxed{2\sqrt{30} \approx 10.95 \text{ rolls}}\end{aligned}$$

Source : https://www.educator.com/media/qn-dl/probability-murray/Prob-14.pdf

The Negative Binomial Distribution

The negative binomial rv and distribution are based on an experiment satisfying the following conditions:

1) The experiment consists of a sequence of independent trials.
2) Each trial can result in either a success (S) or a failure (F).
3) The probability of success is constant from trial to trial, so for $i = 1, 2, 3, \ldots$.
4) The experiment continues (trials are performed) until a total of r successes have been observed, where r is a specified positive integer.

The random variable of interest is X = the number of failures that precede the r^{th} success; X is called a **negative binomial random variable** because, in contrast to the binomial rv, the number of successes is fixed and the number of trials is random.

Source : http://www.auburn.edu/~carpedm/courses/stat3610b/CourseNotesPowerPoint/DevStat8e_03_05.ppt

The Negative Binomial Distribution

Possible values of X are 0, 1, 2, ... Let $nb(x; r, p)$ denote the pmf of X.

Consider $nb(7; 3, p) = P(X = 7)$, the probability that exactly 7 F's occur before the 3rd S.

In order for this to happen, the 10th trial must be an S and there must be exactly 2 S's among the first 9 trials. Thus

$$nb(7; 3, p) = \left\{ \binom{9}{2} \cdot p^2(1-p)^7 \right\} \cdot p = \binom{9}{2} \cdot p^3(1-p)^7$$

Generalizing this line of reasoning gives the following formula for the negative binomial pmf.

Proposition

The pmf of the negative binomial rv X with parameters r = number of S's and $p = P(S)$ is

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x \quad x = 0, 1, 2, \ldots$$

Source : http://www.auburn.edu/~carpedm/courses/stat3610b/CourseNotesPowerPoint/DevStat8e_03_05.ppt

Example

- A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen. Let $p = P$(a randomly selected couple agrees to participate).

- If $p = .2$, what is the probability that 15 couples must be asked before 5 are found who agree to participate? That is, with S = {agrees to participate}, what is the probability that 10 F's occur before the fifth S?

Solution:

Substituting $r = 5$, $p = .2$, and $x = 10$ into $nb(x; r, p)$ gives

$$nb(10; 5, .2) = \binom{14}{4}(.2)^5(.8)^{10} = .034$$

Source : http://www.auburn.edu/~carpedm/courses/stat3610b/CourseNotesPowerPoint/DevStat8e_03_05.ppt

Example Contd.

- The probability that at most 10 *F*'s are observed (at most 15 couples are asked) is

$$P(X \le 10) = \sum_{x=0}^{10} nb(x; 5, .2)$$

$$= (.2)^5 \sum_{x=0}^{10} \binom{x+4}{4} (.8)^x$$

$$= .164$$

Source : http://www.auburn.edu/~carpedm/courses/stat3610b/CourseNotesPowerPoint/DevStat8e_03_05.ppt