Lesson 2.4

The Geometric and Negative Binomial Distributions

Learning Outcomes

At the end of the lesson, students must be able to

1. Describe the geometric and negative binomial distributions,
2. Derive the probability mass functions of a random variable with a geometric or negative binomial distribution,
3. Compute probabilities associated with a random variable with a geometric or negative binomial distribution, and
4. Compute the mean and variance of a random variable with a geometric or negative binomial distribution.

Introduction

The random variable with the geometric probability distribution is associated with an experiment that shares some of the characteristics of a binomial experiment. This experiment also involves identical and independent trials, each of which can result in one of two outcomes: *success* or *failure*. The probability of success is equal to p and is constant from trial to trial. However, instead of the number of successes that occur in a fixed number of trials (n), the geometric random variable Y is the number of the trial on which the first success occurs. Thus, the experiment consists of a series of trials that concludes with the first success. Consequently, the experiment could end with the first trial if a success is observed on the very first trial, or the experiment could go on indefinitely. In other words, the number of trials is not fixed in a geometric distribution.

For example, a person took the Civil Service Examination (CSE) and failed in the first two attempts. He/she wants to know the chance that he/she will pass on the third attempt.

$$P(Y=3) = P(\underbrace{FF}_{\text{2 failures}}\ S) = \underbrace{qq}_{\text{2 terms}}\ p = q^2p$$

Definition:

A random variable Y is said to have a geometric probability distribution with success probability $0 \leq p \leq 1$ if and only

$$P(Y=y) = q^{y-1}p, \text{ where: } y = 1, 2, 3, \cdots,$$

and we write $Y \sim Geom(p)$.

Example 1:

Suppose that 30% of the applicants for a certain industrial job possess advanced training in computer programming. Applicants are interviewed sequentially and are selected at random from the pool. Find the probability that the successful applicant is the fifth one interviewed.

SOLUTION:

Let Y be the interview sequence number of applicants with advanced training in computer programming. Based on the given information, we have $Y \sim Geom(0.3)$. Thus,

$$P(Y=5) = 0.7^4 \times 0.3 = 0.07203$$

Theorem:

If Y has a geometric distribution with success probability p, then

a. $m_Y(t) = \frac{pe^t}{1-qe^t}$

b. $E(Y) = \frac{1}{p}$

c. $V(Y) = \frac{q}{p^2}$

Example 2:

A certified public accountant (CPA) has found that nine of ten company audits contain substantial errors. If the CPA audits a series of company accounts,

a. what is the probability that the first account containing substantial errors is the third one to be audited?

b. what are the mean and standard deviation of the number of accounts?

SOLUTION:

Let Y be the sequence number of the accounts which contains errors. We have $Y \sim Geom(0.9)$.

a. $P(Y = y) = 0.1^2 \times 0.9 = 0.009$

b. The mean is $E(Y) = \frac{1}{0.9} \approx 1$. This average value means that it is very likely that the first account audited contains errors. The standard deviation is $\sigma_Y = \sqrt{\frac{0.1}{0.9^2}} \approx 0.3513$.

Now suppose NEDA R08 has opening for three development specialists. Many new BS Economics graduates applied for the job. The final interview is ongoing. What is the probability that the sixth interviewee is the third to be hired?

Let H and F denote "hired" and "not hired" respectively. One possible sequence of H and F is $FHFHFH$. Other possible sequences of H and F are: $FFHHFH$, $HFFHFH$, and $HHFFFH$. How many possible sequences are there?

There are $\binom{5}{2} = 10$ possible sequences. Under the assumptions of the Bernoulli trials, notice that each sequence has a probability equal to p^3q^3.

Notice that in this example, we needed $Y = 6$ interviews before we can fill in all $r = 3$ vacancies.

This leads us to the following definition.

Definition:

Suppose Bernoulli trials are continually observed. Let Y denote the number of trials required to observe $r \geq 1$ successes. Then

$$P(Y = y) = \binom{y-1}{r-1} p^r q^{y-r}, \quad y = r, r+1, \cdots$$

This is the probability mass function of the negative binomial distribution with success probability equal to p. We write $Y \sim NB(r, p)$.

Remark: When $r = 1$, the negative binomial distribution reduces to the geometric distribution.

Example 3:

Suppose in the motivational example above (NEDA example), the probability that an interviewee is hired is 0.45.

a. What is the probability that the sixth interviewee is the third to be hired?

b. What is the probability that 15 applicants need to be interviewed before all 3 vacancies are filled in?

SOLUTION:

Let Y denote the number of interviews needed to fill in the 3 vacancies, $Y \sim NB(3, 0.45)$.

a. $Y = 6,\ p = 0.45,\ r = 3$

$$P(Y = 6) = \binom{5}{2} \times 0.45^3 \times 0.55^3 \approx 0.1516$$

b. $Y = 15,\ p = 0.45,\ r = 3$

$$P(Y = 15) = \binom{14}{2} \times 0.45^3 \times 0.55^{12} \approx 0.0064$$

Theorem:

If $Y \sim NB(r, p)$, then

a. $m_Y(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$.

b. $E(Y) = \frac{r}{p}$ and $V(Y) = \frac{rq}{p^2}$.

Example 4:

Bob is a high school basketball player. He is a 70% free throw shooter. That means his probability of making a free throw is 0.70.

a. What is the probability that Bob makes his third free throw on his fifth shot?

b. What is the probability his first made free throw is on the third shot?

c. What is the probability it takes more than 3 shots to get his first made free throw?

d. On average, how many attempts does Bob need to make 5 free throws?

SOLUTION: Left as a classroom exercise!