Stat 121 (Mathematical Statistics I)

Lesson 2.5: The Geometric and Negative Binomial Distribution

Learning Outcomes

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

1. Describe random variables that have a 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.

The Geometric Distribution

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 n trials, 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.

The sample space Ω for the experiment contains the countably infinite set of sample points:

$$
\begin{aligned}
E_1 &: \ S, \ \rightarrow \text{success occurs on the first trial} \\
E_2 &: \ FS, \ \rightarrow \text{success occurs on the second trial} \\
E_3 &: \ FFS, \ \rightarrow \text{success occurs on the third trial} \\
&\vdots \\
E_k &: \ \underbrace{FFF\cdots F}_{k-1} \ S, \ \rightarrow \text{success occurs on the kth trial} \\
&\vdots
\end{aligned}
$$

Because the random variable Y is the number of trials up to and including the first success, the events $\{Y = 1\}$, $\{Y = 2\}$, and $\{Y = 3\}$ contain only the sample points $E_1, E_2, andE_3$, respectively. More generally, the numerical event $\{Y = y\}$ contains only E_y. Because the trials are independent, then for any $y = 1, 2, 3, \cdots$

$$P(Y = y) = P(\{\underbrace{FFF\cdots F}_{y-1} \ S\}) = \underbrace{qqq\cdots q}_{y-1} \ p$$

Definition

A random variable Y is said to have a geometric probability distribution with success probability p if and only

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

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

Example 2.5.1

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 first applicant with advanced training in programming is found on the fifth interview.

SOLUTION

Let Y be the interview sequence number where the first applicant with advanced training in programming is found. Based on the given information, we have $Y \sim Geom(0.3)$. Since the first applicant with advanced training in programming is found on the fifth interview, then $Y = 5$, thus,

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

In R you can use the code: **dgeom(4,0.3)** to get this probability

Theorem

If Y has a geometric distribution with success probability p, then its moment generating function is given by

$$m_Y(t) = \frac{pe^t}{1 - qet}$$

PROOF

$$\begin{aligned}
m_Y(t) = E[e^{tY}] &= \sum_{y=1}^{\infty} e^{ty}\, P(Y = y) \\
&= \sum_{y=1}^{\infty} e^{ty} \left[q^{y-1} p \right] \\
&= \frac{p}{q} \sum_{y=1}^{\infty} \left(qe^t \right)^y \\
&= \frac{p}{q} \left(\frac{qe^t}{1 - qe^t} \right) \\
&= \frac{pe^t}{1 - qe^t}
\end{aligned}$$

Theorem

If Y has a geometric distribution with success probability p, then its mean and variance are

$$\begin{aligned}
E(Y) &= \frac{1}{p} \\
V(Y) &= \frac{q}{p^2}
\end{aligned}$$

PROOF: Left as a classroom exercise.

Example 2.5.2

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 that must be examined to find the first one with substantial errors?

SOLUTION: Left as a classroom exercise.

The Negative Binomial Distribution

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.

<u>Example 2.5.3</u>:

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

The moment generating function of the negative binomial distribution with waiting parameter r and success probability p is given by

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

PROOF

Using the definition of the moment generating function and the probability mass function of $Y \sim NB(r, p)$, we have

$$\begin{aligned}
m_Y(t) = E[e^{tY}] &= \sum_{y=r}^{\infty} e^{ty} P(Y = y) \\
&= \sum_{y=r}^{\infty} e^{ty} \binom{y-1}{r-1} p^r q^{y-r} \\
&= \sum_{y=r}^{\infty} e^{ty} \binom{y-1}{r-1} p^r q^{y-r} \ \left[(e^t)^r (e^t)^{-r}\right] \\
&= (pe^t)^r \sum_{y=r}^{\infty} e^{ty} \binom{y-1}{r-1} q^{y-r} (e^t)^{-r} \\
&= (pe^t)^r \sum_{y=r}^{\infty} \binom{y-1}{r-1} q^{y-r} (e^t)^{y-r} \\
&= (pe^t)^r \sum_{y=r}^{\infty} \binom{y-1}{r-1} (qe^t)^{y-r} \\
&= (pe^t)^r (1 - qet)^{-r} \\
&= \frac{(pe^t)^r}{(1 - qet)^r} \\
&= \left(\frac{pe^t}{1 - qet}\right)^r
\end{aligned}$$

Theorem

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

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

SOLUTION: Left as an exercise!

Example 2.5.4:

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!

Learning Task

Instruction: Answer the following as indicated.

1. A door-to-door encyclopedia salesperson is required to document five in-home visits each day. Suppose that she has a 30% chance of being invited into any given home, with each address representing an independent trial. What is the probability that she requires fewer than eight houses to achieve her fifth success?
2. At an automotive plant, 15 percent of all paint batches sent to the lab for chemical analysis do not conform to specifications.

a. What is the probability the third nonconforming batch is observed on the tenth batch sent to the lab?
b. What is the probability that the 6th batch sent to the lab is the first nonconforming batch?