Stat 121 (Mathematical Statistics I)

Lesson 2.4: The Binomial Distribution

Learning Outcomes

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

1. Describe the necessary conditions where a random variable can be modeled using a binomial distribution,
2. Derive the probability mass function of a binomial random variable,
3. Compute probabilities associated with a binomial random variable, and
4. Compute the mean and variance of a binomial random variable.

Introduction

Some experiments consist of the observation of a sequence of identical and independent trials, each of which can result in one of two outcomes. For example, an education graduate may *pass* or *fail* in the teacher's board examination, an item leaving a production line could be *defective* or *not defective*, the sex of a newborn can either be a *boy* or *girl*, and each of a random sample of 35 persons interviewed can be in *favor* or *not (favor)* to a new law.

In general, many experiments consist of a sequence of "trials," where:

1. each trial results in either a "success" or a "failure"
2. the probability of "success," denoted by $p, 0 \leq p \leq 1$, is the same on every trial, and
3. the trials are mutually independent.

Trials that obey these three properties are called **Bernoulli** trials.

In a sequence of Bernoulli trials, one is more interested in the total number of successes. The probability of observing exactly k successes in n independent Bernoulli trials yields the binomial probability distribution. In practice, the binomial probability distribution is used when, for example in a survey of 56 students, we may be interested in the number of them who has the propensity to make online transactions.

Definition:

A *binomial experiment* is one that has the following properties:

1. The experiment consists of n identical (Bernoulli) trials.
2. Each trial results in one of the two outcomes, called a success S and failure F.
3. The probability of success on a single trial is equal to p which remains the same from trial to trial. The probability of failure is $1 - p = q$.
4. The outcomes of the trials are independent.
5. The random variable Y is the number of successes in n trials.

Let us now derive the probability mass function of a binomial random variable. It can be derived by applying the sample point approach to find the probability that the experiment yields y successes. Each sample point in the sample space can be characterized by an n-tuple involving the letters S and F, corresponding to success and failure. A typical sample point would thus appear as

$$\underbrace{SSFSFSSSFF\cdots SF}_{\text{n positions}}$$

where the letter in the i^{th} position (proceeding from left to right) indicates the outcome of the i^{th} trial.

Now let us consider a particular sample point corresponding to y successes and hence contained in the numerical event $Y = y$. This sample point,

$$\underbrace{SSSS\cdots S}_{y}\ \underbrace{FFFF\cdots F}_{n-y}$$

represents the intersection of n independent events (the outcomes of the n trials), in which there were y successes followed by $(n - y)$ failures. Because the trials were independent and the probability of recording a success (S), p, stays the same from trial to trial, the probability of this sample point is

$$\underbrace{pppp\cdots p}_{y}\ \underbrace{qqqq\cdots q}_{n-y} = p^n\, q^{n-y}$$

where $q = 1 - p$is the probability of recording a failure (F).

Note that every other sample point in the event $Y = y$ can be represented as an n-tuple (ordered list) containing y S's and (n - y) F's in some order. Any such sample point also has probability $p^n\, q^{n-y}$. Because the number of n-tuples that contain y S's and $(n - y)$ F's is

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

it follows that the event $(Y = y)$ is made up of $\binom{n}{y}$ sample points, each with probability $p^n\, q^{n-y}$. Thus,

$$P(Y = y) = \binom{n}{y} p^n\, q^{n-y}$$

Definition:

A random variable Y is said to have binomial probability distribution with parameters (n, p) if and only if

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \ y = 0, 1, 2, \cdots, n$$

If Y has a binomial distribution with parameters n and p, we write $Y \sim B(n, p)$.

Example 2.4.1

Upon analyzing the cash register receipts of a large department store over an extended period of time, it is found that 30 percent of the customers pay for their purchases by credit card, 50 percent pay by cash, and 20 percent pay by check. Of the next five customers that make purchases at the store, what is the probability that

a. three of them will pay by credit card?

b. less than 2 will pay by cash?

c. at least 4 will pay by check?

SOLUTION

a. Let X be the number of customers who pay by credit card. We have $n = 5, \ p = 0.30,$ and $y = 3$. Hence,

$$\begin{aligned} P(X = 3) &= \binom{5}{3} (0.3)^3 (1 - 0.3)^{5-3} \\ &= 0.1323 \end{aligned}$$

If you have access to the **R** programming language you can generate the answer using the code: **dbinom(3, 5, 0.3)**.

b. Let Y be the number of customers who pay by cash. This time we have $n = 5, \ p = 0.50,$ and $y = 0$ or 1. Thus,

$$\begin{aligned} P(Y < 2) &= P(Y = 0) + P(Y = 1) \\ &= \binom{5}{0} (0.5)^0 (1 - 0.5)^{5-0} + \binom{5}{1} (0.5)^1 (1 - 0.5)^{5-1} \\ &= 0.1875 \end{aligned}$$

In R you can use either **dbinom(0,5,0.5) + dbinom(1,5,0.5)** or **pbinom(1, 5, 0.5, lower.tail=T)** to get the answer.

c. Let Z be the number of customers who pay by check. In this case $n = 5$, $p = 0.20$, and $y = 4$ or 5. Thus,

$$\begin{aligned} P(Z \geq 4) &= P(Y = 4) + P(Y = 5) \\ &= \binom{5}{4}(0.2)^4(1-0.2)^{5-4} + \binom{5}{5}(0.2)^5(1-0.2)^{5-5} \\ &= 0.00672 \end{aligned}$$

In R you can use either **dbinom(4,5,0.2) + dbinom(5,5,0.2)**, **pbinom(3, 5, 0.2, lower.tail=F)**, or **1 - pbinom(3,5,0.2)** to get the answer.

Example 2.4.2

A new surgical procedure is successful with a probability p. Assume that the operation is performed five times and the results are independent of one another. What is the probability that

a. all five operations are successful if $p = 0.8$?

b. exactly four are successful if $p = 0.6$?

c. less than two are successful if $p = 0.3$?

SOLUTION: Left as a classroom exercise!

Theorem

If $Y \sim B(n, p)$, then its moment generating function is given by

$$m_Y(t) = (pe^t + q)^n$$

Proof

By definition,

$$\begin{aligned} m_Y(t) = E\left[e^{tY}\right] &= \sum_{y=0}^{n} e^{ty} P(Y = y) \\ &= \sum_{y=0}^{n} e^{ty} \binom{n}{y} p^n\, q^{n-y} \\ &= \sum_{y=0}^{n} \binom{n}{y} (pe^t)^y\, q^{n-y} \\ &= (pe^t + q)^n \end{aligned}$$

The last line of the proof was based on the binomial expansion which states that

$$\sum_{r=0}^{n} \binom{n}{r} a^r\, b^{n-r} = (a+b)^n$$

Theorem

If $Y \sim B(n,p)$, then its mean and variance are, respectively, given by

$$E(Y) = np$$
$$V(Y) = npq$$

Proof

Using the definition of the expected value, we have

$$\begin{aligned}
E(Y) &= \sum_{\forall y} yP(Y=y) \\
&= \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y} \\
&= \sum_{y=0}^{n} y \left[\frac{n!}{y(y-1)!(n-y)!}\right] p^y (1-p)^{n-y} \\
&= \sum_{y=0}^{n} \left[\frac{n!}{(y-1)!(n-y)!}\right] p^y (1-p)^{n-y}
\end{aligned}$$

Let $k = y - 1$. This means that $y = k + 1$. If $y = 0$, then $k = -1$, but it is not possible because we are counting the number of successes whose lowest possible value is zero. Hence, the summation starts at $k = 0$. Also, if $y = n$, then $k = n - 1$. Continuing the proof, we have

$$\begin{aligned}
E(Y) &= \sum_{k=0}^{n-1} \left[\frac{n!}{k![n-(k+1)]!}\right] p^{k+1} (1-p)^{n-(k+1)} \\
&= np \sum_{k=0}^{n-1} \left[\frac{(n-1)!}{k![(n-1)-k]!}\right] p^k (1-p)^{(n-1)-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} \\
&= np
\end{aligned}$$

Since by binomial expansion, $\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} = (p+q)^{n-1} = 1$.

Example 2.4.3

Binge drinking is defined as having five or more drinks for male students and four or more drinks for female students at one drinking occasion. Suppose it was found out that the past two weeks approximately 40% of students are engaged in binge drinking. A random sample of 12 students were taken and interviewed.

1. What is the probability that

   a. exactly seven binge drink?
   b. at least 10 binge drink?
   c. at most 4 binge drink?

2. Calculate the mean and standard deviation of the number of students who binge drink.

SOLUTION: Left as a classroom exercise.

Larning Tasks

Instruction: Answer the following as indicated.

1. The probability that a patient recovers from a stomach disease is 0.8. Suppose 20 people are known to have contracted this disease. What is the probability that

   a. exactly 14 recover?
   b. at least 10 recover?
   c. at least 14 but not more than 18 recover?
   d. at most 16 recover?

2. Physicians conjecture that 35 percent of lung cancer patients will respond positively to a new drug treatment. A small clinical trial tests the new drug in 30 patients.

   a. What is the probability that exactly 10 patients respond positively to the drug?
   b. What is the probability that at most 10 patients respond positively to the drug?
   c. What are the mean and variance of the number of patients who respond positively to the drug?

3. Using the definition of expected value, derive the variance of a binomial random variable.