Lesson 2.3: The Binomial Distribution

Norberto E. Milla, Jr.

March 10, 2025

Learning Outcomes

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.

1 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 firm could be defective or not defective, the sex of a newborn can either be a boy or girl, and each of n persons interviewed can be in favor or not to a new law.

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

(i) each trial results in either a "success" or a "failure"

(ii) the probability of "success," denoted by p, $0 \leq p \leq 1$, is the same on every trial, and

(iii) 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 we are concerned with the occurrence of an event, not its magnitude. For example, in a clinical trial, we may be more interested in the number of survivors after a treatment.

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 $q = 1 - p$.
4. The outcomes of the trials are independent.
5. The random variable Y is the number of successes in n trials.

Definition:
A random variable Y is said to have binomial probability distribution with parameters (n, p), written as $binom(n, p)$, if and only if

$$p_Y(y) = P(Y = y) = \binom{n}{y} p^y (1 - p)^{n-y}$$

Example 1:

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:
Let Ybe the number of successful operations. Then $Y \sim binom(5, p)$.

(a) $P(Y = 5|p = 0.8) = \binom{5}{5}0.8^5(1 - 0.8)^{5-5} \approx 0.328$

(b) $P(Y = 4|p = 0.6) = \binom{5}{4}0.6^4(1 - 0.6)^{5-4} \approx 0.259$

(c) $P(Y < 2|p = 0.3) = P(Y = 1) + P(Y = 0) = \binom{5}{1}0.3^1(1-0.3)^{5-1} + \binom{5}{0}0.3^0(1-0.3)^{5-0} \approx 0.528$

Example 2:

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: Left as a classroom exercise!

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

$$m_Y(t) = (q + pe^t)^n, \; q = 1 - p$$

Proof: Left as a classroom exercise!

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

$$\begin{aligned} E(Y) &= np \\ V(Y) &= np(1-p) \end{aligned}$$

Proof: Left as a classroom exercise!

Example 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 students binge drink?
   (b) at least 10 students binge drink?
   (c) at most 4 students binge drink?
2. Calculate the mean and standard deviation of the number of students who binge drink.

Solution: Left as a classroom exercise!