

5.3 Binomial and Multinomial Distributions

An experiment often consists of repeated trials, each with two possible outcomes that may be labeled **success** or **failure**. The most obvious application deals with the testing of items as they come off an assembly line, where each test or trial may indicate a defective or a nondefective item. We may choose to define either outcome as a success. The process is referred to as a **Bernoulli process**. Each trial is called a **Bernoulli trial**. Observe, for example, if one were drawing cards from a deck, the probabilities for repeated trials change if the cards are not replaced. That is, the probability of selecting a heart on the first draw is $1/4$, but on the second draw it is a conditional probability having a value of $13/51$ or $12/51$, depending on whether a heart appeared on the first draw: this, then, would no longer be considered a set of Bernoulli trials.

The Bernoulli Process

Strictly speaking, the Bernoulli process must possess the following properties:

1. The experiment consists of n repeated trials.
2. Each trial results in an outcome that may be classified as a success or a failure.
3. The probability of success, denoted by p , remains constant from trial to trial.
4. The repeated trials are independent.

Consider the set of Bernoulli trials where three items are selected at random from a manufacturing process, inspected, and classified as defective or nondefective. A defective item is designated a success. The number of successes is a random variable X assuming integral values from zero through 3. The eight possible outcomes and the corresponding values of X are

by $\binom{n}{x}$.

Binomial Distribution A Bernoulli trial can result in a success with probability p and a failure with probability $q = 1 - p$. Then the probability distribution of the binomial random variable X , the number of successes in n independent trials, is

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

Note that when $n = 3$ and $p = 1/4$, the probability distribution of X , the number of defectives, may be written as

$$b\left(x; 3, \frac{1}{4}\right) = \binom{3}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{3-x}, \quad x = 0, 1, 2, 3,$$

rather than in the tabular form on page 144.

Example 5.4: The probability that a certain kind of component will survive a shock test is $3/4$. Find the probability that exactly 2 of the next 4 components tested survive.

Solution: Assuming that the tests are independent and $p = 3/4$ for each of the 4 tests, we obtain

$$b\left(2; 4, \frac{3}{4}\right) = \binom{4}{2} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^2 = \left(\frac{4!}{2!2!}\right) \left(\frac{3^2}{4^4}\right) = \frac{27}{128}.$$

Where Does the Name *Binomial* Come From?

The binomial distribution derives its name from the fact that the $n + 1$ terms in the binomial expansion of $(q + p)^n$ correspond to the various values of $b(x; n, p)$ for $x = 0, 1, 2, \dots, n$. That is,

$$\begin{aligned} (q + p)^n &= \binom{n}{0} q^n + \binom{n}{1} p q^{n-1} + \binom{n}{2} p^2 q^{n-2} + \dots + \binom{n}{n} p^n \\ &= b(0; n, p) + b(1; n, p) + b(2; n, p) + \dots + b(n; n, p). \end{aligned}$$

Since $p + q = 1$, we see that

$$\sum_{x=0}^n b(x; n, p) = 1,$$

a condition that must hold for any probability distribution.

Frequently, we are interested in problems where it is necessary to find $P(X < r)$ or $P(a \leq X \leq b)$. Fortunately, binomial sums

$$B(r; n, p) = \sum_{x=0}^r b(x; n, p)$$

are available and are given Table A.1 of the Appendix for $n = 1, 2, \dots, 20$, and selected values of p from 0.1 to 0.9. We illustrate the use of Table A.1 with the following example.

Example 5.5: The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) at least 10 survive, (b) from 3 to 8 survive, and (c) exactly 5 survive?

Solution: Let X be the number of people that survive.

$$(a) \quad P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^9 b(x; 15, 0.4) = 1 - 0.9662 \\ = 0.0338$$

$$(b) \quad P(3 \leq X \leq 8) = \sum_{x=3}^8 b(x; 15, 0.4) = \sum_{x=0}^8 b(x; 15, 0.4) - \sum_{x=0}^2 b(x; 15, 0.4) \\ = 0.9050 - 0.0271 = 0.8779$$

$$(c) \quad P(X = 5) = b(5; 15, 0.4) = \sum_{x=0}^5 b(x; 15, 0.4) - \sum_{x=0}^4 b(x; 15, 0.4) \\ = 0.4032 - 0.2173 = 0.1859$$

Example 5.6: A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%.

- (a) The inspector of the retailer randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?
- (b) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be 3 shipments containing at least one defective device?

Solution: (a) Denote by X the number of defective devices among the 20. This X follows a $b(x; 20, 0.03)$ distribution. Hence

$$P(X \geq 1) = 1 - P(X = 0) = 1 - b(0; 20, 0.03) \\ = 1 - 0.03^0(1 - 0.03)^{20-0} = 0.4562.$$

- (b) In this case, each shipment can either contain at least one defective item or not. Hence, testing the result of each shipment can be viewed as a Bernoulli trial with $p = 0.4562$ from part (a). Assuming the independence from shipment to shipment and denoting by Y the number of shipments containing at least one defective item Y follows another binomial distribution $b(y; 10, 0.4562)$. Therefore, the answer to this question is

$$P(Y = 3) = \binom{10}{3} 0.4562^3 (1 - 0.4562)^7 = 0.1602.$$

of recoveries among 15 patients subjected to the given disease has a probability of at least $3/4$ of falling between 2.206 and 9.794, or, because the data are discrete, between 3 and 9 inclusive.

5.8: It is conjectured that an impurity exists in 30% of all drinking wells in a certain rural community. In order to gain some insight on this problem, it is determined that some tests should be made. It is too expensive to test all of the many wells in the area, so 10 were randomly selected for testing.

(a) Using the binomial distribution, what is the probability that exactly three wells have the impurity assuming that the conjecture is correct?

(b) What is the probability that more than three wells are impure?

tion: (a) We require

$$\begin{aligned} b(3; 10, 0.3) = P(X = 3) &= \sum_{x=0}^3 b(x; 10, 0.3) - \sum_{x=0}^2 b(x; 10, 0.3) \\ &= 0.6496 - 0.3828 = 0.2668. \end{aligned}$$

(b) In this case we need $P(X > 3) = 1 - 0.6496 = 0.3504$.

There are solutions in which the computation of binomial probabilities may allow us to draw inference about a scientific population after data are collected. An illustration is given in the next example.

5.9: Consider the situation of Example 5.8. The "30% are impure" is merely a conjecture put forth by the area water board. Suppose 10 wells are randomly selected and 6 are found to contain the impurity. What does this imply about the conjecture? Use a probability statement.

tion: We must first ask: "If the conjecture is correct, is it likely that we could have found 6 or more impure wells?"

$$P(X \geq 6) = \sum_{x=6}^{10} b(x; 10, 0.3) = 1 - \sum_{x=0}^5 b(x; 10, 0.3) = 1 - 0.9527 = 0.0473.$$

As a result, it is very unlikely (4.7% chance) that 6 or more wells would be found impure if only 30% of all are impure. This casts considerable doubt on the conjecture and suggests that the impurity problem is much more severe.

As the reader should realize by now, in many applications there are more than two possible outcomes. To borrow an example from the field of genetics, the color of guinea pigs produced as offspring may be red, black, or white. Often the "defective" or "not defective" dichotomy in engineering situations is truly an oversimplification. Indeed, there are often more than two categories that characterize items or parts coming off an assembly line.

22: In a certain industrial facility accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other.

(a) What is the probability that in any given period of 400 days there will be an accident on one day?

(b) What is the probability that there are at most three days with an accident?

on: Let X be a binomial random variable with $n = 400$ and $p = 0.005$. Thus $np = 2$. Using the Poisson approximation,

(a) $P(X = 1) = e^{-2}2^1 = 0.271$, and

(b) $P(X \leq 3) = \sum_{x=0}^3 e^{-2}2^x/x! = 0.857$.

3: In a manufacturing process where glass products are produced, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

on: This is essentially a binomial experiment with $n = 8000$ and $p = 0.001$. Since p is very close to zero and n is quite large, we shall approximate with the Poisson distribution using

$$\mu = (8000)(0.001) = 8.$$

Hence, if X represents the number of bubbles, we have

$$P(X < 7) = \sum_{x=0}^6 b(x; 8000, 0.001) \approx p(x; 8) = 0.3134.$$