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

Outcome	x
NNN	0
NDN	1
NND	1
DNN	1
NDD	2
DND	2
DDN	2
DDD	3

Since the items are selected independently from a process that we shall assume produces 25% defectives,

$$P(NDN) = P(N)P(D)P(N) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{9}{64}.$$

Similar calculations yield the probabilities for the other possible outcomes. The probability distribution of X is therefore

x	0	1	2	3
$f(x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

The number X of successes in n Bernoulli trials is called a **binomial random variable**. The probability distribution of this discrete random variable is called the **binomial distribution**, and its values will be denoted by $b(x; n, p)$ since they depend on the number of trials and the probability of a success on a given trial. Thus, for the probability distribution of X, the number of defectives is

$$P(X = 2) = f(2) = b\left(2; 3, \frac{1}{4}\right) = \frac{9}{64}.$$

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

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 $x = 0, 1, 2, \ldots, n$. That is,

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

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 $n = 1, 2, \ldots, 20$, 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) $$P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662$$
$$= 0.0338$$

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

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

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

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

Theorem 5.2: The mean and variance of the binomial distribution $b(x; n, p)$ are

$$\mu = np, \quad \text{and} \quad \sigma^2 = npq.$$

***Proof*:** Let the outcome on the jth trial be represented by a Bernoulli random variable I_j, which assumes the values 0 and 1 with probabilities q and p, respectively. Therefore, in a binomial experiment the number of successes can be written as the sum of the n independent indicator variables. Hence

$$X = I_1 + I_2 + \cdots + I_n.$$

The mean of any I_j is $E(I_j) = (0)(q) + (1)(p) = p$. Therefore, using Corollary 4.4, the mean of the binomial distribution is

$$\mu = E(X) = E(I_1) + E(I_2) + \cdots + E(I_n) = \underbrace{p + p + \cdots + p}_{n \text{ terms}} = np.$$

The variance of any I_j is

$$\sigma_{I_j}^2 = E[(I_j - p)^2] = E(I_j^2) - p^2 = (0)^2(q) + (1)^2(p) - p^2 = p(1-p) = pq.$$

By extending Corollary 4.10 to the case of n independent variables, the variance of the binomial distribution is

$$\sigma_X^2 = \sigma_{I_1}^2 + \sigma_{I_2}^2 + \cdots + \sigma_{I_n}^2 = \underbrace{pq + pq + \cdots + pq}_{n \text{ terms}} = npq.$$

Example 5.7: Find the mean and variance of the binomial random variable of Example 5.5, and then use Chebyshev's theorem (on page 132) to interpret the interval $\mu \pm 2\sigma$.

***Solution*:** Since Example 5.5 was a binomial experiment with $n = 15$ and $p = 0.4$, by Theorem 5.2, we have

$$\mu = (15)(0.4) = 6, \quad \text{and} \quad \sigma^2 = (15)(0.4)(0.6) = 3.6.$$

Taking the square root of 3.6, we find that $\sigma = 1.897$. Hence the required interval is $6 \pm (2)(1.897)$, or from 2.206 to 9.794. Chebyshev's theorem states that the number 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.

Example 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?

Solution: (a) We require

$$b(3; 10, 0.3) = P(X = 3) = \sum_{x=0}^{3} b(x; 10, 0.3) - \sum_{x=0}^{2} b(x; 10, 0.3)$$

$$= 0.6496 - 0.3828 = 0.2668.$$

(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.

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.

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

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

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.