# Chapter 5

# Some Discrete Probability Distributions

## 5.1 Introduction and Motivation

No matter whether a discrete probability distribution is represented graphically by a histogram, in tabular form, or by means of a formula, the behavior of a random variable is described. Often, the observations generated by different statistical experiments have the same general type of behavior. Consequently, discrete random variables associated with these experiments can be described by essentially the same probability distribution and therefore can be represented by a single formula. In fact, one needs only a handful of important probability distributions to describe many of the discrete random variables encountered in practice.

Such a handful of distributions describe several real-life random phenomena. For instance, in a study involving testing the effectiveness of a new drug, the number of cured patients among all the patients who use the drug approximately follows a binomial distribution (Section 5.2). In an industrial example, when a sample of items selected from a batch of production is tested, the number of defective items in the sample usually can be modeled as a hypergeometric random variable (Section 5.3). In a statistical quality control problem, the experimenter will signal a shift of the process mean when observational data exceed certain limits. The number of samples required to produce a false alarm follows a geometric distribution which is a special case of the negative binomial distribution (Section 5.4). On the other hand, the number of white cells from a fixed amount of an individual's blood sample is usually random and may be described by a Poisson distribution (Section 5.5). In this chapter, we present these commonly used distributions with various examples.

## 5.2 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 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 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 0 through 3. The eight possible outcomes and the corresponding values of $X$ are

| **Outcome** | $NNN$ | $NDN$ | $NND$ | $DNN$ | $NDD$ | $DND$ | $DDN$ | $DDD$ |
|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{x}$ | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 3 |

Since the items are selected independently and we assume that the process produces 25% defectives, we have

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

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

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

## Binomial Distribution

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

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

Let us now generalize the above illustration to yield a formula for $b(x; n, p)$. That is, we wish to find a formula that gives the probability of $x$ successes in $n$ trials for a binomial experiment. First, consider the probability of $x$ successes and $n - x$ failures in a specified order. Since the trials are independent, we can multiply all the probabilities corresponding to the different outcomes. Each success occurs with probability $p$ and each failure with probability $q = 1 - p$. Therefore, the probability for the specified order is $p^x q^{n-x}$. We must now determine the total number of sample points in the experiment that have $x$ successes and $n - x$ failures. This number is equal to the number of partitions of $n$ outcomes into two groups with $x$ in one group and $n - x$ in the other and is written $\binom{n}{x}$ as introduced in Section 2.3. Because these partitions are mutually exclusive, we add the probabilities of all the different partitions to obtain the general formula, or simply multiply $p^x q^{n-x}$ 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

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

---

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 5.1:** 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,

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

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 $P(a \leq X \leq b)$. Binomial sums

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

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

**Example 5.2:** 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 who 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 5.3:** 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 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 exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

***Solution:*** (a) Denote by $X$ the number of defective devices among the 20. Then $X$ 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 of each shipment can be viewed as a Bernoulli trial with $p = 0.4562$ from part (a). Assuming 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,

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

## Areas of Application

From Examples 5.1 through 5.3, it should be clear that the binomial distribution finds applications in many scientific fields. An industrial engineer is keenly interested in the "proportion defective" in an industrial process. Often, quality control measures and sampling schemes for processes are based on the binomial distribution. This distribution applies to any industrial situation where an outcome of a process is dichotomous and the results of the process are independent, with the probability of success being constant from trial to trial. The binomial distribution is also used extensively for medical and military applications. In both fields, a success or failure result is important. For example, "cure" or "no cure" is important in pharmaceutical work, and "hit" or "miss" is often the interpretation of the result of firing a guided missile.

Since the probability distribution of any binomial random variable depends only on the values assumed by the parameters $n$, $p$, and $q$, it would seem reasonable to assume that the mean and variance of a binomial random variable also depend on the values assumed by these parameters. Indeed, this is true, and in the proof of Theorem 5.1 we derive general formulas that can be used to compute the mean and variance of any binomial random variable as functions of $n$, $p$, and $q$.

**Theorem 5.1:** The mean and variance of the binomial distribution $b(x; n, p)$ are

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

***Proof*:** Let the outcome on the $j$th 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 on page 131, 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 $\sigma^2_{I_j} = E(I_j^2) - p^2 = (0)^2(q) + (1)^2(p) - p^2 = p(1-p) = pq$. Extending Corollary 4.11 to the case of $n$ independent Bernoulli variables gives the variance of the binomial distribution as

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

**Example 5.4:** It is conjectured that an impurity exists in 30% of all drinking wells in a certain rural community. In order to gain some insight into the true extent of the problem, it is determined that some testing is necessary. It is too expensive to test all of the wells in the area, so 10 are randomly selected for testing.

(a) Using the binomial distribution, what is the probability that exactly 3 wells have the impurity, assuming that the conjecture is correct?

(b) What is the probability that more than 3 wells are impure?

***Solution*:** (a) We require

$$b(3; 10, 0.3) = \sum_{x=0}^{3} b(x; 10, 0.3) - \sum_{x=0}^{2} b(x; 10, 0.3) = 0.6496 - 0.3828 = 0.2668.$$

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

**Example 5.5:** Find the mean and variance of the binomial random variable of Example 5.2, and then use Chebyshev's theorem (on page 137) to interpret the interval $\mu \pm 2\sigma$.

***Solution*:** Since Example 5.2 was a binomial experiment with $n = 15$ and $p = 0.4$, by Theorem 5.1, we have

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

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 who contracted the disease has a probability of at least 3/4 of falling between 2.206 and 9.794 or, because the data are discrete, between 2 and 10 inclusive.

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

**Example 5.6:** Consider the situation of Example 5.4. The notion that 30% of the wells 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 would find 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 is truly an oversimplification in engineering situations. Indeed, there are often more than two categories that characterize items or parts coming off an assembly line.

## Multinomial Experiments and the Multinomial Distribution

The binomial experiment becomes a **multinomial experiment** if we let each trial have more than two possible outcomes. The classification of a manufactured product as being light, heavy, or acceptable and the recording of accidents at a certain intersection according to the day of the week constitute multinomial experiments. The drawing of a card from a deck *with replacement* is also a multinomial experiment if the 4 suits are the outcomes of interest.

In general, if a given trial can result in any one of $k$ possible outcomes $E_1, E_2, \ldots, E_k$ with probabilities $p_1, p_2, \ldots, p_k$, then the **multinomial distribution** will give the probability that $E_1$ occurs $x_1$ times, $E_2$ occurs $x_2$ times, ..., and $E_k$ occurs $x_k$ times in $n$ independent trials, where

$$x_1 + x_2 + \cdots + x_k = n.$$

We shall denote this joint probability distribution by

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n).$$

Clearly, $p_1 + p_2 + \cdots + p_k = 1$, since the result of each trial must be one of the $k$ possible outcomes.

To derive the general formula, we proceed as in the binomial case. Since the trials are independent, any specified order yielding $x_1$ outcomes for $E_1$, $x_2$ for $E_2, \ldots, x_k$ for $E_k$ will occur with probability $p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$. The total number of orders yielding similar outcomes for the $n$ trials is equal to the number of partitions of $n$ items into $k$ groups with $x_1$ in the first group, $x_2$ in the second group, ..., and $x_k$ in the $k$th group. This can be done in

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1!\, x_2! \cdots x_k!}$$

ways. Since all the partitions are mutually exclusive and occur with equal probability, we obtain the multinomial distribution by multiplying the probability for a specified order by the total number of partitions.

---

**Multinomial Distribution**

If a given trial can result in the $k$ outcomes $E_1, E_2, \ldots, E_k$ with probabilities $p_1, p_2, \ldots, p_k$, then the probability distribution of the random variables $X_1, X_2, \ldots, X_k$, representing the number of occurrences for $E_1, E_2, \ldots, E_k$ in $n$ independent trials, is

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k},$$

with

$$\sum_{i=1}^{k} x_i = n \text{ and } \sum_{i=1}^{k} p_i = 1.$$

---

The multinomial distribution derives its name from the fact that the terms of the multinomial expansion of $(p_1 + p_2 + \cdots + p_k)^n$ correspond to all the possible values of $f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n)$.

**Example 5.7:** The complexity of arrivals and departures of planes at an airport is such that computer simulation is often used to model the "ideal" conditions. For a certain airport with three runways, it is known that in the ideal setting the following are the probabilities that the individual runways are accessed by a randomly arriving commercial jet:

Runway 1: $p_1 = 2/9$,
Runway 2: $p_2 = 1/6$,
Runway 3: $p_3 = 11/18$.

What is the probability that 6 randomly arriving airplanes are distributed in the following fashion?

Runway 1: 2 airplanes,
Runway 2: 1 airplane,
Runway 3: 3 airplanes

***Solution*:** Using the multinomial distribution, we have

$$f\left(2, 1, 3; \frac{2}{9}, \frac{1}{6}, \frac{11}{18}, 6\right) = \binom{6}{2, 1, 3}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3$$
$$= \frac{6!}{2!\,1!\,3!} \cdot \frac{2^2}{9^2} \cdot \frac{1}{6} \cdot \frac{11^3}{18^3} = 0.1127.$$

## Exercises

**5.1** A random variable $X$ that assumes the values $x_1, x_2, \ldots, x_k$ is called a discrete uniform random variable if its probability mass function is $f(x) = \frac{1}{k}$ for all of $x_1, x_2, \ldots, x_k$ and 0 otherwise. Find the mean and variance of $X$.

**5.2** Twelve people are given two identical speakers, which they are asked to listen to for differences, if any. Suppose that these people answer simply by guessing. Find the probability that three people claim to have heard a difference between the two speakers.

**5.3** An employee is selected from a staff of 10 to supervise a certain project by selecting a tag at random from a box containing 10 tags numbered from 1 to 10. Find the formula for the probability distribution of $X$ representing the number on the tag that is drawn. What is the probability that the number drawn is less than 4?

**5.4** In a certain city district, the need for money to buy drugs is stated as the reason for 75% of all thefts. Find the probability that among the next 5 theft cases reported in this district,

(a) exactly 2 resulted from the need for money to buy drugs;

(b) at most 3 resulted from the need for money to buy drugs.

**5.5** According to *Chemical Engineering Progress* (November 1990), approximately 30% of all pipework failures in chemical plants are caused by operator error.

(a) What is the probability that out of the next 20 pipework failures at least 10 are due to operator error?

(b) What is the probability that no more than 4 out of 20 such failures are due to operator error?

(c) Suppose, for a particular plant, that out of the random sample of 20 such failures, exactly 5 are due to operator error. Do you feel that the 30% figure stated above applies to this plant? Comment.

**5.6** According to a survey by the Administrative Management Society, one-half of U.S. companies give employees 4 weeks of vacation after they have been with the company for 15 years. Find the probability that among 6 companies surveyed at random, the number that give employees 4 weeks of vacation after 15 years of employment is

(a) anywhere from 2 to 5;

(b) fewer than 3.

**5.7** One prominent physician claims that 70% of those with lung cancer are chain smokers. If his assertion is correct,

(a) find the probability that of 10 such patients recently admitted to a hospital, fewer than half are chain smokers;

(b) find the probability that of 20 such patients recently admitted to a hospital, fewer than half are chain smokers.

**5.8** According to a study published by a group of University of Massachusetts sociologists, approximately 60% of the Valium users in the state of Massachusetts first took Valium for psychological problems. Find the probability that among the next 8 users from this state who are interviewed,

(a) exactly 3 began taking Valium for psychological problems;

(b) at least 5 began taking Valium for problems that were not psychological.

**5.9** In testing a certain kind of truck tire over rugged terrain, it is found that 25% of the trucks fail to complete the test run without a blowout. Of the next 15 trucks tested, find the probability that

(a) from 3 to 6 have blowouts;

(b) fewer than 4 have blowouts;

(c) more than 5 have blowouts.

**5.10** A nationwide survey of college seniors by the University of Michigan revealed that almost 70% disapprove of daily pot smoking, according to a report in *Parade*. If 12 seniors are selected at random and asked their opinion, find the probability that the number who disapprove of smoking pot daily is

(a) anywhere from 7 to 9;

(b) at most 5;

(c) not less than 8.

**5.11** The probability that a patient recovers from a delicate heart operation is 0.9. What is the probability that exactly 5 of the next 7 patients having this operation survive?

**5.12** A traffic control engineer reports that 75% of the vehicles passing through a checkpoint are from within the state. What is the probability that fewer than 4 of the next 9 vehicles are from out of state?

**5.13** A national study that examined attitudes about antidepressants revealed that approximately 70% of respondents believe "antidepressants do not really cure anything, they just cover up the real trouble." According to this study, what is the probability that at least 3 of the next 5 people selected at random will hold this opinion?

**5.14** The percentage of wins for the Chicago Bulls basketball team going into the playoffs for the 1996–97 season was 87.7. Round the 87.7 to 90 in order to use Table A.1.

(a) What is the probability that the Bulls sweep (4-0) the initial best-of-7 playoff series?

(b) What is the probability that the Bulls win the initial best-of-7 playoff series?

(c) What very important assumption is made in answering parts (a) and (b)?

**5.15** It is known that 60% of mice inoculated with a serum are protected from a certain disease. If 5 mice are inoculated, find the probability that

(a) none contracts the disease;

(b) fewer than 2 contract the disease;

(c) more than 3 contract the disease.

**5.16** Suppose that airplane engines operate independently and fail with probability equal to 0.4. Assuming that a plane makes a safe flight if at least one-half of its engines run, determine whether a 4-engine plane or a 2-engine plane has the higher probability for a successful flight.

**5.17** If $X$ represents the number of people in Exercise 5.13 who believe that antidepressants do not cure but only cover up the real problem, find the mean and variance of $X$ when 5 people are selected at random.

**5.18** (a) In Exercise 5.9, how many of the 15 trucks would you expect to have blowouts?

(b) What is the variance of the number of blowouts experienced by the 15 trucks? What does that mean?

**5.19** As a student drives to school, he encounters a traffic signal. This traffic signal stays green for 35 seconds, yellow for 5 seconds, and red for 60 seconds. Assume that the student goes to school each weekday between 8:00 and 8:30 a.m. Let $X_1$ be the number of times he encounters a green light, $X_2$ be the number of times he encounters a yellow light, and $X_3$ be the number of times he encounters a red light. Find the joint distribution of $X_1$, $X_2$, and $X_3$.

**5.20** According to *USA Today* (March 18, 1997), of 4 million workers in the general workforce, 5.8% tested positive for drugs. Of those testing positive, 22.5% were cocaine users and 54.4% marijuana users.

(a) What is the probability that of 10 workers testing positive, 2 are cocaine users, 5 are marijuana users, and 3 are users of other drugs?

(b) What is the probability that of 10 workers testing positive, all are marijuana users?

(c) What is the probability that of 10 workers testing positive, none is a cocaine user?

**5.21** The surface of a circular dart board has a small center circle called the bull's-eye and 20 pie-shaped regions numbered from 1 to 20. Each of the pie-shaped regions is further divided into three parts such that a person throwing a dart that lands in a specific region scores the value of the number, double the number, or triple the number, depending on which of the three parts the dart hits. If a person hits the bull's-eye with probability 0.01, hits a double with probability 0.10, hits a triple with probability 0.05, and misses the dart board with probability 0.02, what is the probability that 7 throws will result in no bull's-eyes, no triples, a double twice, and a complete miss once?

**5.22** According to a genetics theory, a certain cross of guinea pigs will result in red, black, and white offspring in the ratio 8:4:4. Find the probability that among 8 offspring, 5 will be red, 2 black, and 1 white.

**5.23** The probabilities are 0.4, 0.2, 0.3, and 0.1, respectively, that a delegate to a certain convention arrived by air, bus, automobile, or train. What is the probability that among 9 delegates randomly selected at this convention, 3 arrived by air, 3 arrived by bus, 1 arrived by automobile, and 2 arrived by train?

**5.24** A safety engineer claims that only 40% of all workers wear safety helmets when they eat lunch at the workplace. Assuming that this claim is right, find the probability that 4 of 6 workers randomly chosen will be wearing their helmets while having lunch at the workplace.

**5.25** Suppose that for a very large shipment of integrated-circuit chips, the probability of failure for any one chip is 0.10. Assuming that the assumptions underlying the binomial distributions are met, find the probability that at most 3 chips fail in a random sample of 20.

**5.26** Assuming that 6 in 10 automobile accidents are due mainly to a speed violation, find the probability that among 8 automobile accidents, 6 will be due mainly to a speed violation

(a) by using the formula for the binomial distribution;

(b) by using Table A.1.

**5.27** If the probability that a fluorescent light has a useful life of at least 800 hours is 0.9, find the probabilities that among 20 such lights

(a) exactly 18 will have a useful life of at least 800 hours;

(b) at least 15 will have a useful life of at least 800 hours;

(c) at least 2 will *not* have a useful life of at least 800 hours.

**5.28** A manufacturer knows that on average 20% of the electric toasters produced require repairs within 1 year after they are sold. When 20 toasters are randomly selected, find appropriate numbers $x$ and $y$ such that

(a) the probability that at least $x$ of them will require repairs is less than 0.5;

(b) the probability that at least $y$ of them will *not* require repairs is greater than 0.8.

## 5.3 Hypergeometric Distribution

The simplest way to view the distinction between the binomial distribution of Section 5.2 and the hypergeometric distribution is to note the way the sampling is done. The types of applications for the hypergeometric are very similar to those for the binomial distribution. We are interested in computing probabilities for the number of observations that fall into a particular category. But in the case of the binomial distribution, independence among trials is required. As a result, if that distribution is applied to, say, sampling from a lot of items (deck of cards, batch of production items), the sampling must be done **with replacement** of each item after it is observed. On the other hand, the hypergeometric distribution does not require independence and is based on sampling done **without replacement**.

Applications for the hypergeometric distribution are found in many areas, with heavy use in acceptance sampling, electronic testing, and quality assurance. Obviously, in many of these fields, testing is done at the expense of the item being tested. That is, the item is destroyed and hence cannot be replaced in the sample. Thus, sampling without replacement is necessary. A simple example with playing cards will serve as our first illustration.

If we wish to find the probability of observing 3 red cards in 5 draws from an ordinary deck of 52 playing cards, the binomial distribution of Section 5.2 does not apply unless each card is replaced and the deck reshuffled before the next draw is made. To solve the problem of sampling without replacement, let us restate the problem. If 5 cards are drawn at random, we are interested in the probability of selecting 3 red cards from the 26 available in the deck and 2 black cards from the 26 available in the deck. There are $\binom{26}{3}$ ways of selecting 3 red cards, and for each of these ways we can choose 2 black cards in $\binom{26}{2}$ ways. Therefore, the total number of ways to select 3 red and 2 black cards in 5 draws is the product $\binom{26}{3}\binom{26}{2}$. The total number of ways to select any 5 cards from the 52 that are available is $\binom{52}{5}$. Hence, the probability of selecting 5 cards without replacement of which 3 are red and 2 are black is given by

$$\frac{\binom{26}{3}\binom{26}{2}}{\binom{52}{5}} = \frac{(26!/3!\,23!)(26!/2!\,24!)}{52!/5!\,47!} = 0.3251.$$

In general, we are interested in the probability of selecting $x$ successes from the $k$ items labeled successes and $n - x$ failures from the $N - k$ items labeled failures when a random sample of size $n$ is selected from $N$ items. This is known as a **hypergeometric experiment**, that is, one that possesses the following two properties:

1. A random sample of size $n$ is selected without replacement from $N$ items.
2. Of the $N$ items, $k$ may be classified as successes and $N - k$ are classified as failures.

The number $X$ of successes of a hypergeometric experiment is called a **hypergeometric random variable**. Accordingly, the probability distribution of the hypergeometric variable is called the **hypergeometric distribution**, and its values are denoted by $h(x; N, n, k)$, since they depend on the number of successes $k$ in the set $N$ from which we select $n$ items.

## Hypergeometric Distribution in Acceptance Sampling

Like the binomial distribution, the hypergeometric distribution finds applications in acceptance sampling, where lots of materials or parts are sampled in order to determine whether or not the entire lot is accepted.

**Example 5.8:** A particular part that is used as an injection device is sold in lots of 10. The producer deems a lot acceptable if no more than one defective is in the lot. A sampling plan involves random sampling and testing 3 of the parts out of 10. If none of the 3 is defective, the lot is accepted. Comment on the utility of this plan.

***Solution*:** Let us assume that the lot is truly **unacceptable** (i.e., that 2 out of 10 parts are defective). The probability that the sampling plan finds the lot acceptable is

$$P(X = 0) = \frac{\binom{2}{0}\binom{8}{3}}{\binom{10}{3}} = 0.467.$$

Thus, if the lot is truly unacceptable, with 2 defective parts, this sampling plan will allow acceptance roughly 47% of the time. As a result, this plan should be considered faulty.

Let us now generalize in order to find a formula for $h(x; N, n, k)$. The total number of samples of size $n$ chosen from $N$ items is $\binom{N}{n}$. These samples are assumed to be equally likely. There are $\binom{k}{x}$ ways of selecting $x$ successes from the $k$ that are available, and for each of these ways we can choose the $n-x$ failures in $\binom{N-k}{n-x}$ ways. Thus, the total number of favorable samples among the $\binom{N}{n}$ possible samples is given by $\binom{k}{x}\binom{N-k}{n-x}$. Hence, we have the following definition.

**Hypergeometric Distribution**

The probability distribution of the hypergeometric random variable $X$, the number of successes in a random sample of size $n$ selected from $N$ items of which $k$ are labeled **success** and $N-k$ labeled **failure**, is

$$h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \quad \max\{0, n-(N-k)\} \leq x \leq \min\{n, k\}.$$

The range of $x$ can be determined by the three binomial coefficients in the definition, where $x$ and $n-x$ are no more than $k$ and $N-k$, respectively, and both of them cannot be less than 0. Usually, when both $k$ (the number of successes) and $N-k$ (the number of failures) are larger than the sample size $n$, the range of a hypergeometric random variable will be $x = 0, 1, \ldots, n$.

**Example 5.9:** Lots of 40 components each are deemed unacceptable if they contain 3 or more defectives. The procedure for sampling a lot is to select 5 components at random and to reject the lot if a defective is found. What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot?

***Solution*:** Using the hypergeometric distribution with $n = 5$, $N = 40$, $k = 3$, and $x = 1$, we find the probability of obtaining 1 defective to be

$$h(1; 40, 5, 3) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011.$$

Once again, this plan is not desirable since it detects a bad lot (3 defectives) only about 30% of the time.

**Theorem 5.2:** The mean and variance of the hypergeometric distribution $h(x; N, n, k)$ are

$$\mu = \frac{nk}{N} \text{ and } \sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right).$$

The proof for the mean is shown in Appendix A.24.

**Example 5.10:** Let us now reinvestigate Example 3.4 on page 83. The purpose of this example was to illustrate the notion of a random variable and the corresponding sample space. In the example, we have a lot of 100 items of which 12 are defective. What is the probability that in a sample of 10, 3 are defective?

***Solution*:** Using the hypergeometric probability function, we have

$$h(3; 100, 10, 12) = \frac{\binom{12}{3}\binom{88}{7}}{\binom{100}{10}} = 0.08.$$

**Example 5.11:** Find the mean and variance of the random variable of Example 5.9 and then use Chebyshev's theorem to interpret the interval $\mu \pm 2\sigma$.

***Solution*:** Since Example 5.9 was a hypergeometric experiment with $N = 40$, $n = 5$, and $k = 3$, by Theorem 5.2, we have

$$\mu = \frac{(5)(3)}{40} = \frac{3}{8} = 0.375,$$

and

$$\sigma^2 = \left(\frac{40-5}{39}\right)(5)\left(\frac{3}{40}\right)\left(1 - \frac{3}{40}\right) = 0.3113.$$

Taking the square root of 0.3113, we find that $\sigma = 0.558$. Hence, the required interval is $0.375 \pm (2)(0.558)$, or from $-0.741$ to $1.491$. Chebyshev's theorem states that the number of defectives obtained when 5 components are selected at random from a lot of 40 components of which 3 are defective has a probability of at least 3/4 of falling between $-0.741$ and $1.491$. That is, at least three-fourths of the time, the 5 components include fewer than 2 defectives.

## Relationship to the Binomial Distribution

In this chapter, we discuss several important discrete distributions that have wide applicability. Many of these distributions relate nicely to each other. The beginning student should gain a clear understanding of these relationships. There is an interesting relationship between the hypergeometric and the binomial distribution. As one might expect, if $n$ is small compared to $N$, the nature of the $N$ items changes very little in each draw. So a binomial distribution can be used to approximate the hypergeometric distribution when $n$ is small compared to $N$. In fact, as a rule of thumb, the approximation is good when $n/N \leq 0.05$.

Thus, the quantity $k/N$ plays the role of the binomial parameter $p$. As a result, the binomial distribution may be viewed as a large-population version of the hypergeometric distribution. The mean and variance then come from the formulas

$$\mu = np = \frac{nk}{N} \text{ and } \sigma^2 = npq = n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right).$$

Comparing these formulas with those of Theorem 5.2, we see that the mean is the same but the variance differs by a correction factor of $(N - n)/(N - 1)$, which is negligible when $n$ is small relative to $N$.

**Example 5.12:** A manufacturer of automobile tires reports that among a shipment of 5000 sent to a local distributor, 1000 are slightly blemished. If one purchases 10 of these tires at random from the distributor, what is the probability that exactly 3 are blemished?

***Solution:*** Since $N = 5000$ is large relative to the sample size $n = 10$, we shall approximate the desired probability by using the binomial distribution. The probability of obtaining a blemished tire is 0.2. Therefore, the probability of obtaining exactly 3 blemished tires is

$$h(3; 5000, 10, 1000) \approx b(3; 10, 0.2) = 0.8791 - 0.6778 = 0.2013.$$

On the other hand, the exact probability is $h(3; 5000, 10, 1000) = 0.2015$.

The hypergeometric distribution can be extended to treat the case where the $N$ items can be partitioned into $k$ cells $A_1, A_2, \ldots, A_k$ with $a_1$ elements in the first cell, $a_2$ elements in the second cell, ..., $a_k$ elements in the $k$th cell. We are now interested in the probability that a random sample of size $n$ yields $x_1$ elements from $A_1$, $x_2$ elements from $A_2$, ..., and $x_k$ elements from $A_k$. Let us represent this probability by

$$f(x_1, x_2, \ldots, x_k; a_1, a_2, \ldots, a_k, N, n).$$

To obtain a general formula, we note that the total number of samples of size $n$ that can be chosen from $N$ items is still $\binom{N}{n}$. There are $\binom{a_1}{x_1}$ ways of selecting $x_1$ items from the items in $A_1$, and for each of these we can choose $x_2$ items from the items in $A_2$ in $\binom{a_2}{x_2}$ ways. Therefore, we can select $x_1$ items from $A_1$ and $x_2$ items from $A_2$ in $\binom{a_1}{x_1}\binom{a_2}{x_2}$ ways. Continuing in this way, we can select all $n$ items consisting of $x_1$ from $A_1$, $x_2$ from $A_2$, ..., and $x_k$ from $A_k$ in

$$\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k} \quad \text{ways.}$$

The required probability distribution is now defined as follows.

**Multivariate Hypergeometric Distribution**

If $N$ items can be partitioned into the $k$ cells $A_1, A_2, \ldots, A_k$ with $a_1, a_2, \ldots, a_k$ elements, respectively, then the probability distribution of the random variables $X_1, X_2, \ldots, X_k$, representing the number of elements selected from $A_1, A_2, \ldots, A_k$ in a random sample of size $n$, is

$$f(x_1, x_2, \ldots, x_k; a_1, a_2, \ldots, a_k, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k}}{\binom{N}{n}},$$

with $\sum_{i=1}^{k} x_i = n$ and $\sum_{i=1}^{k} a_i = N$.

**Example 5.13:** A group of 10 individuals is used for a biological case study. The group contains 3 people with blood type O, 4 with blood type A, and 3 with blood type B. What is the probability that a random sample of 5 will contain 1 person with blood type O, 2 people with blood type A, and 2 people with blood type B?

***Solution:*** Using the extension of the hypergeometric distribution with $x_1 = 1$, $x_2 = 2$, $x_3 = 2$, $a_1 = 3$, $a_2 = 4$, $a_3 = 3$, $N = 10$, and $n = 5$, we find that the desired probability is

$$f(1, 2, 2; 3, 4, 3, 10, 5) = \frac{\binom{3}{1}\binom{4}{2}\binom{3}{2}}{\binom{10}{5}} = \frac{3}{14}.$$

## Exercises

**5.29** A homeowner plants 6 bulbs selected at random from a box containing 5 tulip bulbs and 4 daffodil bulbs. What is the probability that he planted 2 daffodil bulbs and 4 tulip bulbs?

**5.30** To avoid detection at customs, a traveler places 6 narcotic tablets in a bottle containing 9 vitamin tablets that are similar in appearance. If the customs official selects 3 of the tablets at random for analysis, what is the probability that the traveler will be arrested for illegal possession of narcotics?

**5.31** A random committee of size 3 is selected from 4 doctors and 2 nurses. Write a formula for the probability distribution of the random variable $X$ representing the number of doctors on the committee. Find $P(2 \leq X \leq 3)$.

**5.32** From a lot of 10 missiles, 4 are selected at random and fired. If the lot contains 3 defective missiles that will not fire, what is the probability that

(a) all 4 will fire?

(b) at most 2 will not fire?

**5.33** If 7 cards are dealt from an ordinary deck of 52 playing cards, what is the probability that

(a) exactly 2 of them will be face cards?

(b) at least 1 of them will be a queen?

**5.34** What is the probability that a waitress will refuse to serve alcoholic beverages to only 2 minors if she randomly checks the IDs of 5 among 9 students, 4 of whom are minors?

**5.35** A company is interested in evaluating its current inspection procedure for shipments of 50 identical items. The procedure is to take a sample of 5 and pass the shipment if no more than 2 are found to be defective. What proportion of shipments with 20% defectives will be accepted?

**5.36** A manufacturing company uses an acceptance scheme on items from a production line before they are shipped. The plan is a two-stage one. Boxes of 25 items are readied for shipment, and a sample of 3 items is tested for defectives. If any defectives are found, the entire box is sent back for 100% screening. If no defectives are found, the box is shipped.

(a) What is the probability that a box containing 3 defectives will be shipped?

(b) What is the probability that a box containing only 1 defective will be sent back for screening?

**5.37** Suppose that the manufacturing company of Exercise 5.36 decides to change its acceptance scheme. Under the new scheme, an inspector takes 1 item at random, inspects it, and then replaces it in the box; a second inspector does likewise. Finally, a third inspector goes through the same procedure. The box is not shipped if any of the three inspectors find a defective. Answer the questions in Exercise 5.36 for this new plan.

**5.38** Among 150 IRS employees in a large city, only 30 are women. If 10 of the employees are chosen at random to provide free tax assistance for the residents of this city, use the binomial approximation to the hypergeometric distribution to find the probability that at least 3 women are selected.

**5.39** An annexation suit against a county subdivision of 1200 residences is being considered by a neighboring city. If the occupants of half the residences object to being annexed, what is the probability that in a random sample of 10 at least 3 favor the annexation suit?

**5.40** It is estimated that 4000 of the 10,000 voting residents of a town are against a new sales tax. If 15 eligible voters are selected at random and asked their opinion, what is the probability that at most 7 favor the new tax?

**5.41** A nationwide survey of 17,000 college seniors by the University of Michigan revealed that almost 70% disapprove of daily pot smoking. If 18 of these seniors are selected at random and asked their opinion, what is the probability that more than 9 but fewer than 14 disapprove of smoking pot daily?

**5.42** Find the probability of being dealt a bridge hand of 13 cards containing 5 spades, 2 hearts, 3 diamonds, and 3 clubs.

**5.43** A foreign student club lists as its members 2 Canadians, 3 Japanese, 5 Italians, and 2 Germans. If a committee of 4 is selected at random, find the probability that

(a) all nationalities are represented;

(b) all nationalities except Italian are represented.

**5.44** An urn contains 3 green balls, 2 blue balls, and 4 red balls. In a random sample of 5 balls, find the probability that both blue balls and at least 1 red ball are selected.

**5.45** Biologists doing studies in a particular environment often tag and release subjects in order to estimate the size of a population or the prevalence of certain features in the population. Ten animals of a certain population thought to be extinct (or near extinction) are caught, tagged, and released in a certain region. After a period of time, a random sample of 15 of this type of animal is selected in the region. What is the probability that 5 of those selected are tagged if there are 25 animals of this type in the region?

**5.46** A large company has an inspection system for the batches of small compressors purchased from vendors. A batch typically contains 15 compressors. In the inspection system, a random sample of 5 is selected and all are tested. Suppose there are 2 faulty compressors in the batch of 15.

(a) What is the probability that for a given sample there will be 1 faulty compressor?

(b) What is the probability that inspection will discover both faulty compressors?

**5.47** A government task force suspects that some manufacturing companies are in violation of federal pollution regulations with regard to dumping a certain type of product. Twenty firms are under suspicion but not all can be inspected. Suppose that 3 of the firms are in violation.

(a) What is the probability that inspection of 5 firms will find no violations?

(b) What is the probability that the plan above will find two violations?

**5.48** Every hour, 10,000 cans of soda are filled by a machine, among which 300 underfilled cans are produced. Each hour, a sample of 30 cans is randomly selected and the number of ounces of soda per can is checked. Denote by $X$ the number of cans selected that are underfilled. Find the probability that at least 1 underfilled can will be among those sampled.

## 5.4 Negative Binomial and Geometric Distributions

Let us consider an experiment where the properties are the same as those listed for a binomial experiment, with the exception that the trials will be repeated until a *fixed* number of successes occur. Therefore, instead of the probability of $x$ successes in $n$ trials, where $n$ is fixed, we are now interested in the probability that the $k$th success occurs on the $x$th trial. Experiments of this kind are called **negative binomial experiments**.

As an illustration, consider the use of a drug that is known to be effective in 60% of the cases where it is used. The drug will be considered a success if it is effective in bringing some degree of relief to the patient. We are interested in finding the probability that the fifth patient to experience relief is the seventh patient to receive the drug during a given week. Designating a success by $S$ and a failure by $F$, a possible order of achieving the desired result is $SFSSSFS$, which occurs with probability

$$(0.6)(0.4)(0.6)(0.6)(0.6)(0.4)(0.6) = (0.6)^5(0.4)^2.$$

We could list all possible orders by rearranging the $F$'s and $S$'s except for the last outcome, which must be the fifth success. The total number of possible orders is equal to the number of partitions of the first six trials into two groups with 2 failures assigned to the one group and 4 successes assigned to the other group. This can be done in $\binom{6}{4} = 15$ mutually exclusive ways. Hence, if $X$ represents the outcome on which the fifth success occurs, then

$$P(X = 7) = \binom{6}{4}(0.6)^5(0.4)^2 = 0.1866.$$

### What Is the Negative Binomial Random Variable?

The number $X$ of trials required to produce $k$ successes in a negative binomial experiment is called a **negative binomial random variable**, and its probability distribution is called the **negative binomial distribution**. Since its probabilities depend on the number of successes desired and the probability of a success on a given trial, we shall denote them by $b^*(x;k,p)$. To obtain the general formula for $b^*(x;k,p)$, consider the probability of a success on the $x$th trial preceded by $k-1$ successes and $x-k$ failures in some specified order. Since the trials are independent, we can multiply all the probabilities corresponding to each desired outcome. Each success occurs with probability $p$ and each failure with probability $q = 1-p$. Therefore, the probability for the specified order ending in success is

$$p^{k-1}q^{x-k}p = p^k q^{x-k}.$$

The total number of sample points in the experiment ending in a success, after the occurrence of $k-1$ successes and $x-k$ failures in any order, is equal to the number of partitions of $x-1$ trials into two groups with $k-1$ successes corresponding to one group and $x-k$ failures corresponding to the other group. This number is specified by the term $\binom{x-1}{k-1}$, each mutually exclusive and occurring with equal probability $p^k q^{x-k}$. We obtain the general formula by multiplying $p^k q^{x-k}$ by $\binom{x-1}{k-1}$.

---

**Negative Binomial Distribution**

If repeated independent trials can result in a success with probability $p$ and a failure with probability $q = 1-p$, then the probability distribution of the random variable $X$, the number of the trial on which the $k$th success occurs, is

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

---

**Example 5.14:** In an NBA (National Basketball Association) championship series, the team that wins four games out of seven is the winner. Suppose that teams $A$ and $B$ face each other in the championship games and that team $A$ has probability 0.55 of winning a game over team $B$.

(a) What is the probability that team $A$ will win the series in 6 games?

(b) What is the probability that team $A$ will win the series?

(c) If teams $A$ and $B$ were facing each other in a regional playoff series, which is decided by winning three out of five games, what is the probability that team $A$ would win the series?

***Solution:*** (a) $b^*(6;4,0.55) = \binom{5}{3} 0.55^4 (1-0.55)^{6-4} = 0.1853$

(b) $P$(team $A$ wins the championship series) is

$$\begin{aligned} &b^*(4;4,0.55) + b^*(5;4,0.55) + b^*(6;4,0.55) + b^*(7;4,0.55) \\ &= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083. \end{aligned}$$

(c) $P$(team $A$ wins the playoff) is

$$\begin{aligned} &b^*(3;3,0.55) + b^*(4;3,0.55) + b^*(5;3,0.55) \\ &= 0.1664 + 0.2246 + 0.2021 = 0.5931. \end{aligned}$$

The negative binomial distribution derives its name from the fact that each term in the expansion of $p^k(1-q)^{-k}$ corresponds to the values of $b^*(x;k,p)$ for $x = k,\ k+1,\ k+2,\ \ldots$. If we consider the special case of the negative binomial distribution where $k = 1$, we have a probability distribution for the number of trials required for a single success. An example would be the tossing of a coin until a head occurs. We might be interested in the probability that the first head occurs on the fourth toss. The negative binomial distribution reduces to the form

$$b^*(x;1,p) = pq^{x-1}, \quad x = 1,2,3,\ldots.$$

Since the successive terms constitute a geometric progression, it is customary to refer to this special case as the **geometric distribution** and denote its values by $g(x;p)$.

**Geometric Distribution** If repeated independent trials can result in a success with probability $p$ and a failure with probability $q = 1-p$, then the probability distribution of the random variable $X$, the number of the trial on which the first success occurs, is

$$g(x;p) = pq^{x-1}, \quad x = 1,2,3,\ldots.$$

**Example 5.15:** For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

***Solution*:** Using the geometric distribution with $x = 5$ and $p = 0.01$, we have

$$g(5;0.01) = (0.01)(0.99)^4 = 0.0096.$$

**Example 5.16:** At a "busy time," a telephone exchange is very near capacity, so callers have difficulty placing their calls. It may be of interest to know the number of attempts necessary in order to make a connection. Suppose that we let $p = 0.05$ be the probability of a connection during a busy time. We are interested in knowing the probability that 5 attempts are necessary for a successful call.

***Solution*:** Using the geometric distribution with $x = 5$ and $p = 0.05$ yields

$$P(X = x) = g(5;0.05) = (0.05)(0.95)^4 = 0.041.$$

Quite often, in applications dealing with the geometric distribution, the mean and variance are important. For example, in Example 5.16, the *expected* number of calls necessary to make a connection is quite important. The following theorem states without proof the mean and variance of the geometric distribution.

**Theorem 5.3:** The mean and variance of a random variable following the geometric distribution are

$$\mu = \frac{1}{p} \text{ and } \sigma^2 = \frac{1-p}{p^2}.$$

## Applications of Negative Binomial and Geometric Distributions

Areas of application for the negative binomial and geometric distributions become obvious when one focuses on the examples in this section and the exercises devoted to these distributions at the end of Section 5.5. In the case of the geometric distribution, Example 5.16 depicts a situation where engineers or managers are attempting to determine how inefficient a telephone exchange system is during busy times. Clearly, in this case, trials occurring prior to a success represent a cost. If there is a high probability of several attempts being required prior to making a connection, then plans should be made to redesign the system.

Applications of the negative binomial distribution are similar in nature. Suppose attempts are costly in some sense and are *occurring in sequence*. A high probability of needing a "large" number of attempts to experience a fixed number of successes is not beneficial to the scientist or engineer. Consider the scenarios of Review Exercises 5.90 and 5.91. In Review Exercise 5.91, the oil driller defines a certain level of success from sequentially drilling locations for oil. If only 6 attempts have been made at the point where the second success is experienced, the profits appear to dominate substantially the investment incurred by the drilling.