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 kth success occurs on the xth 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 xth 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 kth 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

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

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

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

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.

5.5 Poisson Distribution and the Poisson Process

Experiments yielding numerical values of a random variable X, the number of outcomes occurring during a given time interval or in a specified region, are called **Poisson experiments**. The given time interval may be of any length, such as a minute, a day, a week, a month, or even a year. For example, a Poisson experiment can generate observations for the random variable X representing the number of telephone calls received per hour by an office, the number of days school is closed due to snow during the winter, or the number of games postponed due to rain during a baseball season. The specified region could be a line segment, an area, a volume, or perhaps a piece of material. In such instances, X might represent the number of field mice per acre, the number of bacteria in a given culture, or the number of typing errors per page. A Poisson experiment is derived from the **Poisson process** and possesses the following properties.

Properties of the Poisson Process

1. The number of outcomes occurring in one time interval or specified region of space is independent of the number that occur in any other disjoint time interval or region. In this sense we say that the Poisson process has no memory.
2. The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region and does not depend on the number of outcomes occurring outside this time interval or region.
3. The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.

The number X of outcomes occurring during a Poisson experiment is called a **Poisson random variable**, and its probability distribution is called the **Poisson**