1

BASIC CONCEPTS IN PROBABILITY

1.1 INTRODUCTION

The concepts of *experiments* and *events* are very important in the study of probability. In probability, an experiment is any process of trial and observation. An experiment whose outcome is uncertain before it is performed is called a *random* experiment. When we perform a random experiment, the collection of possible elementary outcomes is called the *sample space* of the experiment, which is usually denoted by Ω. We define these outcomes as elementary outcomes because exactly one of the outcomes occurs when the experiment is performed. The elementary outcomes of an experiment are called the *sample points* of the sample space and are denoted by w_i, $i = 1, 2, \ldots$ If there are n possible outcomes of an experiment, then the sample space is $\Omega = \{w_1, w_2, \ldots, w_n\}$. An *event* is the occurrence of either a prescribed outcome or any one of a number of possible outcomes of an experiment. Thus, an event is a subset of the sample space.

1.2 RANDOM VARIABLES

Consider a random experiment with sample space Ω. Let w be a sample point in Ω. We are interested in assigning a real number to each $w \in \Omega$. A random variable, $X(w)$, is a single-valued real function that assigns a real number,

Fundamentals of Stochastic Networks, First Edition. Oliver C. Ibe.

called the value of $X(w)$, to each sample point $w \in \Omega$. That is, it is a mapping of the sample space onto the real line.

Generally a random variable is represented by a single letter X instead of the function $X(w)$. Therefore, in the remainder of the book we use X to denote a random variable. The sample space Ω is called the *domain* of the random variable X. Also, the collection of all numbers that are values of X is called the *range* of the random variable X.

Let X be a random variable and x a fixed real value. Let the event A_x define the subset of Ω that consists of all real sample points to which the random variable X assigns the number x.

That is,

$$A_x = \{w|X(w) = x\} = [X = x].$$

Since A_x is an event, it will have a probability, which we define as follows:

$$p = P[A_x].$$

We can define other types of events in terms of a random variable. For fixed numbers x, a, and b, we can define the following:

$$[X \le x] = \{w|X(w) \le x\},$$
$$[X > x] = \{w|X(w) > x\},$$
$$[a < X < b] = \{w|a < X(w) < b\}.$$

These events have probabilities that are denoted by

- $P[X \le x]$ is the probability that X takes a value less than or equal to x.
- $P[X > x]$ is the probability that X takes a value greater than x; this is equal to $1 - P[X \le x]$.
- $P[a < X < b]$ is the probability that X takes a value that strictly lies between a and b.

1.2.1 Distribution Functions

Let X be a random variable and x be a number. As stated earlier, we can define the event $[X \le x] = \{x|X(w) \le x\}$. The distribution function (or the cumulative distribution function [CDF]) of X is defined by:

$$F_X(x) = P[X \le x] \quad -\infty < x < \infty.$$

That is, $F_X(x)$ denotes the probability that the random variable X takes on a value that is less than or equal to x. Some properties of $F_X(x)$ include:

1. $F_X(x)$ is a nondecreasing function, which means that if $x_1 < x_2$, then $F_X(x_1) \le F_X(x_2)$. Thus, $F_X(x)$ can increase or stay level, but it cannot go down.

2. $0 \leq F_X(x) \leq 1$
3. $F_X(\infty) = 1$
4. $F_X(-\infty) = 0$
5. $P[a < X \leq b] = F_X(b) - F_X(a)$
6. $P[X > a] = 1 - P[X \leq a] = 1 - F_X(a)$

1.2.2 Discrete Random Variables

A discrete random variable is a random variable that can take on at most a countable number of possible values. For a discrete random variable X, the *probability mass function* (PMF), $p_X(x)$, is defined as follows:

$$p_X(x) = P[X = x].$$

The PMF is nonzero for at most a countable or countably infinite number of values of x. In particular, if we assume that X can only assume one of the values $x_1, x_2, \ldots, x_n$, then:

$$\begin{aligned} p_X(x_i) &\geq 0 \quad i = 1, 2 \ldots, n, \\ p_X(x) &= 0 \quad \text{otherwise.} \end{aligned}$$

The CDF of X can be expressed in terms of $p_X(x)$ as follows:

$$F_x(x) = \sum_{k \leq x} p_X(k).$$

The CDF of a discrete random variable is a step function. That is, if X takes on values $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \ldots$, then the value of $F_X(x)$ is constant in the interval between x_{i-1} and x_i and then takes a jump of size $p_X(x_i)$ at x_i, $i = 2, 3, \ldots$. Thus, in this case, $F_X(x)$ represents the sum of all the probability masses we have encountered as we move from $-\infty$ to x.

1.2.3 Continuous Random Variables

Discrete random variables have a set of possible values that are either finite or countably infinite. However, there exists another group of random variables that can assume an uncountable set of possible values. Such random variables are called continuous random variables. Thus, we define a random variable X to be a continuous random variable if there exists a nonnegative function $f_X(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set A of real numbers,

$$P[X \in A] = \int_A f_X(x)\,dx.$$

The function $f_X(x)$ is called the *probability density function* (PDF) of the random variable X and is defined by:

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

The properties of $f_X(x)$ are as follows:

1. $f_X(x) \geq 0$
2. Since X must assume some value, $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
3. $P[a \leq X \leq b] = \int_a^b f_X(x)\,dx$, which means that $P[X = a] = \int_a^a f_X(x)\,dx = 0$. Thus, the probability that a continuous random variable will assume any fixed value is zero.
4. $P[X < a] = P[X \leq a] = F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx$

1.2.4 Expectations

If X is a random variable, then the *expectation* (or *expected value* or *mean*) of X, denoted by E[X], is defined by:

$$E[X] = \begin{cases} \sum_i x_i p_X(x_i) & X \quad \text{discrete} \\ \int_{-\infty}^{\infty} x f_X(x)\,dx & X \quad \text{continuous} \end{cases}$$

Thus, the expected value of X is a weighted average of the possible values that X can take, where each value is weighted by the probability that X takes that value. The expected value of X is sometimes denoted by $\bar{X}$.

1.2.5 Moments of Random Variables and the Variance

The nth moment of the random variable X, denoted by $E[X^n] = \overline{X^n}$, is defined by:

$$E[X^n] = \overline{X^n} = \begin{cases} \sum_i x_i^n p_X(x_i) & X \quad \text{discrete} \\ \int_{-\infty}^{\infty} x^n f_x(x)\,dx & X \quad \text{continuous} \end{cases}$$

for $n = 1, 2, 3, \ldots$. The first moment, $E[X]$, is the expected value of X.

We can also define the *central moments* (or *moments about the mean*) of a random variable. These are the moments of the difference between a random variable and its expected value. The nth central moment is defined by

$$E\left[(X - \bar{X})^n\right] = \overline{(X - \bar{X})^n} = \begin{cases} \sum_i (x_i - \bar{X})^n p_X(x_i) & X \quad \text{discrete} \\ \int_{-\infty}^{\infty} (x - \bar{X})^n f_X(x)\,dx & X \quad \text{continuous} \end{cases}$$

The central moment for the case of $n = 2$ is very important and carries a special name, the *variance*, which is usually denoted by σ_X^2. Thus,

$$\sigma_X^2 = E\left[\left(X-\overline{X}\right)^2\right] = \overline{\left(X-\overline{X}\right)^2} = \begin{cases} \sum_i \left(x_i-\overline{X}\right)^2 p_X(x_i) & X \quad \text{discrete} \\ \int_{-\infty}^{\infty} \left(x-\overline{X}\right)^2 f_X(x)\,dx & X \quad \text{continuous} \end{cases}$$

1.3 TRANSFORM METHODS

Different types of transforms are used in science and engineering. In this book we consider two types of transforms: the z-transform of PMFs and the s-transform of PDFs of nonnegative random variables. These transforms are particularly used when random variables take only nonnegative values, which is usually the case in many applications discussed in this book.

1.3.1 The s-Transform

Let $f_X(x)$ be the PDF of the continuous random variable X that takes only nonnegative values; that is, $f_X(x) = 0$ for $x < 0$. The s-transform of $f_X(x)$, denoted by $M_X(s)$, is defined by:

$$M_X(s) = E\left[e^{-sX}\right] = \int_0^{\infty} e^{-sx} f_X(x)\,dx.$$

One important property of an s-transform is that when it is evaluated at the point $s = 0$, its value is equal to 1. That is,

$$M_X(s)\big|_{s=0} = \int_0^{\infty} f_X(x)\,dx = 1.$$

For example, the value of K for which the function $A(s) = K/(s+5)$ is a valid s-transform of a PDF is obtained by setting $A(0) = 1$, which gives:

$$K/5 = 1 \Rightarrow K = 5.$$

1.3.2 Moment-Generating Property of the s-Transform

One of the primary reasons for studying the transform methods is to use them to derive the moments of the different probability distributions. By definition:

$$M_X(s) = \int_0^{\infty} e^{-sx} f_X(x)\,dx.$$

Taking different derivatives of $M_X(s)$ and evaluating them at $s = 0$, we obtain the following results:

$$
\begin{aligned}
\frac{d}{ds}M_X(s) &= \frac{d}{ds}\int_0^\infty e^{-sx} f_X(x)\,dx = \int_0^\infty \frac{d}{ds} e^{-sx} f_X(x)\,dx \\
&= -\int_0^\infty x e^{-sx} f_X(x)\,dx, \\
\frac{d}{ds}M_X(s)\big|_{s=0} &= -\int_0^\infty x f_X(x)\,dx \\
&= -E[X], \\
\frac{d^2}{ds^2}M_X(s) &= \frac{d}{ds}(-1)\int_{-\infty}^\infty x e^{-sx} f_X(x)\,dx = \int_0^\infty x^2 e^{-sx} f_X(x)\,dx, \\
\frac{d^2}{ds^2}M_X(s)\big|_{s=0} &= \int_0^\infty x^2 f_X(x)\,dx \\
&= E\left[X^2\right].
\end{aligned}
$$

In general,

$$
\frac{d^n}{ds^n}M_X(s)\big|_{s=0} = (-1)^n E\left[X^n\right].
$$

1.3.3 The z-Transform

Let $p_X(x)$ be the PMF of the discrete random variable X. The z-transform of $p_X(x)$, denoted by $G_X(z)$, is defined by:

$$
G_X(z) = E\left[z^X\right] = \sum_{x=0}^{\infty} z^x p_X(x).
$$

Thus, the PMF $p_X(x)$ is required to take on only nonnegative integers, as we stated earlier. The sum is guaranteed to converge and, therefore, the z-transform exists, when evaluated on or within the unit circle (where $|z| \le 1$). Note that:

$$
G_X(1) = \sum_{x=0}^{\infty} p_X(x) = 1.
$$

This means that a valid z-transform of a PMF reduces to unity when evaluated at $z = 1$. However, this is a necessary but not sufficient condition for a function to the z-transform of a PMF. By definition,

$$
\begin{aligned}
G_X(z) &= \sum_{x=0}^{\infty} z^x p_X(x) \\
&= p_X(0) + z p_X(1) + z^2 p_X(2) + z^3 p_X(3) + \dots.
\end{aligned}
$$

This means that $P[X = k] = p_X(k)$ is the coefficient of z^k in the series expansion. Thus, given the z-transform of a PMF, we can uniquely recover the PMF.

The implication of this statement is that not every function of z that has a value of 1 when evaluated at $z = 1$ is a valid z-transform of a PMF. For example, consider the function $A(z) = 2z - 1$. Although $A(1) = 1$, the function contains invalid coefficients in the sense that these coefficients either have negative values or positive values that are greater than one. Thus, for a function of z to be a valid z-transform of a PMF, it must have a value of 1 when evaluated at $z = 1$, and the coefficients of z must be nonnegative numbers that cannot be greater than 1.

The individual terms of the PMF can also be determined as follows:

$$p_X(x) = \frac{1}{x!}\left[\frac{d^x}{dz^x}G_X(z)\right]_{z=0} \qquad x = 0, 1, 2, \ldots.$$

This feature of the z-transform is the reason it is sometimes called the *probability generating function*.

1.3.4 Moment-Generating Property of the z-Transform

As stated earlier, one of the major motivations for studying transform methods is their usefulness in computing the moments of the different random variables. Unfortunately, the moment-generating capability of the z-transform is not as computationally efficient as that of the s-transform.

The moment-generating capability of the z-transform lies in the results obtained from evaluating the derivatives of the transform at $z = 1$. For a discrete random variable X with PMF $p_X(x)$, we have that:

$$G_X(z) = \sum_{x=0}^{\infty} z^x p_X(x),$$

$$\frac{d}{dz}G_X(z) = \frac{d}{dz}\sum_{x=0}^{\infty} z^x p_X(x) = \sum_{x=0}^{\infty}\frac{d}{dz}z^x p_X(x) = \sum_{x=0}^{\infty} xz^{x-1}p_X(x) = \sum_{x=1}^{\infty} xz^{x-1}p_X(x),$$

$$\frac{d}{dz}G_X(z)\Big|_{z=1} = \sum_{x=1}^{\infty} xp_X(x) = \sum_{x=0}^{\infty} xp_X(x) = E[X].$$

Similarly,

$$\frac{d^2}{dz^2}G_X(z) = \frac{d}{dz}\sum_{x=1}^{\infty} xz^{x-1}p_X(x) = \sum_{x=1}^{\infty} x\frac{d}{dz}z^{x-1}p_X(x) = \sum_{x=1}^{\infty} x(x-1)z^{x-2}p_X(x),$$

$$\begin{aligned}\frac{d^2}{dz^2}G_X(z)\Big|_{z=1} &= \sum_{x=1}^{\infty} x(x-1)p_X(x) = \sum_{x=0}^{\infty} x(x-1)p_X(x) = \sum_{x=0}^{\infty} x^2 p_X(x) - \sum_{x=0}^{\infty} xp_X(x)\\ &= E[X^2] - E[X],\end{aligned}$$

$$E[X^2] = \frac{d^2}{dz^2}G_X(z)\Big|_{z=1} + \frac{d}{dz}G_X(z)\Big|_{z=1}.$$

Thus, the variance is obtained as follows:

$$\begin{aligned}\sigma_X^2 &= E[X^2]-(E[X])^2\\ &=\left[\frac{d^2}{dz^2}G_X(z)+\frac{d}{dz}G_X(z)-\left\{\frac{d}{dz}G_X(z)\right\}^2\right]_{z=1}.\end{aligned}$$

1.4 COVARIANCE AND CORRELATION COEFFICIENT

Consider two random variables X and Y with expected values $E[X] = \mu_X$ and $E[Y] = \mu_Y$, respectively, and variances σ_X^2 and σ_Y^2, respectively. The *covariance* of X and Y, which is denoted by $Cov(X, Y)$ or σ_{XY}, is defined by:

$$\begin{aligned}Cov(X,Y)=\sigma_{XY} &= E[(X-\mu_X)(Y-\mu_Y)]\\ &= E[XY-\mu_Y X-\mu_X Y+\mu_X\mu_Y]\\ &= E[XY]-\mu_X\mu_Y-\mu_X\mu_Y+\mu_X\mu_Y\\ &= E[XY]-\mu_X\mu_Y.\end{aligned}$$

If X and Y are independent, then $E[XY] = \mu_X\mu_Y$ and $Cov(X, Y) = 0$. However, the converse is not true; that is, if the covariance of X and Y is zero, it does not mean that X and Y are independent random variables. If the covariance of two random variables is zero, we define the two random variables to be *uncorrelated*.

We define the *correlation coefficient* of X and Y, denoted by $\rho(X, Y)$ or ρ_{XY}, as follows:

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

The correlation coefficient has the property that:

$$-1\le\rho_{XY}\le 1.$$

1.5 SUMS OF INDEPENDENT RANDOM VARIABLES

Consider two independent continuous random variables X and Y. We are interested in computing the CDF and PDF of their sum $g(X, Y) = U = X + Y$. The random variable S can be used to model the reliability of systems with stand-by connections. In such systems, the component A whose time-to-failure is represented by the random variable X is the primary component, and the component B whose time-to-failure is represented by the random variable Y is the backup component that is brought into operation when the primary component fails. Thus, S represents the time until the system fails, which is the sum of the lifetimes of both components.

Their CDF can be obtained as follows:

$$F_S(s) = P[S \le s] = P[X + Y \le s] = \iint_D f_{XY}(x, y)\,dxdy,$$

where $f_{XY}(x,y)$ is the joint PDF of X and Y and D is the set $D = \{(x,y)|x + y \le s\}$. Thus,

$$\begin{aligned} F_S(s) &= \int_{-\infty}^{\infty}\int_{-\infty}^{s-y} f_{XY}(x, y)\,dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{s-y} f_X(x) f_Y(y)\,dxdy \\ &= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{s-y} f_X(x)\,dx\right\} f_Y(y)\,dy \\ &= \int_{-\infty}^{\infty} F_X(s-y) f_Y(y)\,dy. \end{aligned}$$

The PDF of S is obtained by differentiating the CDF, as follows:

$$\begin{aligned} f_S(s) &= \frac{d}{ds} F_S(s) = \frac{d}{ds}\int_{-\infty}^{\infty} F_X(s-y) f_Y(y)\,dy \\ &= \int_{-\infty}^{\infty} \frac{d}{ds} F_X(s-y) f_Y(y)\,dy \\ &= \int_{-\infty}^{\infty} f_X(s-y) f_Y(y)\,dy, \end{aligned}$$

where we have assumed that we can interchange differentiation and integration. The expression on the right-hand side is a well-known result in signal analysis called the *convolution integral*. Thus, we find that the PDF of the sum S of two independent random variables X and Y is the convolution of the PDFs of the two random variables; that is,

$$f_S(s) = f_X(s) * f_Y(s).$$

In general, if S is the sum on n mutually independent random variables X_1, $X_2, \ldots, X_n$ whose PDFs are $f_{X_i}(x)$, $i = 1, 2, \ldots, n$, then we have that:

$$\begin{gathered} S = X_1 + X_2 + \ldots + X_n, \\ f_S(s) = f_{X_1}(s) * f_{X_2}(s) \ldots * f_{X_n}(s). \end{gathered}$$

Thus, the s-transform of the PDF of S is given by:

$$M_S(s) = \prod_{i=1}^{n} M_{X_i}(s).$$

1.6 RANDOM SUM OF RANDOM VARIABLES

Let X be a continuous random variable with PDF $f_X(x)$ whose s-transform is $M_X(s)$. We know that if Y is the sum of n independent and identically distributed random variables with the PDF $f_X(x)$, then from the results in the previous section, the s-transform of the PDF of Y is given by:

$$M_Y(s) = [M_X(s)]^n.$$

This result assumes that n is a fixed number. However, there are certain situations when the number of random variables in a sum is itself a random variable. For this case, let N denote a discrete random variable with PMF $p_N(n)$ whose z-transform is $G_N(z)$. Our goal is to find the s-transform of the PDF of Y when the number of random variables is itself a random variable N.

Thus, we consider the sum:

$$Y = X_1 + X_2 + \ldots + X_N,$$

where N has a known PMF, which in turn has a known z-transform. Now, let $N = n$. Then with N fixed at n, we have that:

$$\begin{aligned}
Y|_{N=n} &= X_1 + X_2 + \ldots + X_n, \\
M_{Y|N}(s|n) &= [M_X(s)]^n, \\
M_Y(s) &= \sum_n p_N(n) M_{Y|N}(s|n) = \sum_n p_N(n)[M_X(s)]^n = G_N(M_X(s)).
\end{aligned}$$

That is, the s-transform of the PDF of a random sum of independent and identically distributed random variables is the z-transform of the PMF of the number of variables evaluated at the s-transform of the PDF of the constituent random variables. Now, let $u = M_X(s)$. Then,

$$\begin{aligned}
\frac{d}{ds}M_Y(s) &= \frac{d}{ds}G_N(M_X(s)) = \left\{\frac{dG_N(u)}{du}\right\}\left\{\frac{du}{ds}\right\}, \\
\frac{d}{ds}M_Y(s)\Big|_{s=0} &= \left[\left\{\frac{dG_N(u)}{du}\right\}\left\{\frac{du}{ds}\right\}\right]_{s=0}.
\end{aligned}$$

When $s = 0$, $u|_{s=0} = M_X(0) = 1$. Thus, we obtain:

$$\begin{aligned}
\frac{d}{ds}M_Y(s)\Big|_{s=0} &= \left[\left\{\frac{dG_N(u)}{du}\right\}\left\{\frac{du}{ds}\right\}\right]_{s=0} = \frac{dG_N(u)}{du}\Big|_{u=1}\frac{dM_X(s)}{ds}\Big|_{s=0}, \\
-E[Y] &= E[N](-E[X]) = -E[N]E[X], \\
E[Y] &= E[N]E[X].
\end{aligned}$$

Also,

$$\begin{aligned}
\frac{d^2}{ds^2}M_Y(s) &= \frac{d}{ds}\left[\left\{\frac{dG_N(u)}{du}\right\}\left\{\frac{du}{ds}\right\}\right] = \left\{\frac{du}{ds}\right\}\frac{d}{ds}\left\{\frac{dG_N(u)}{du}\right\} + \left\{\frac{dG_N(u)}{du}\right\}\left\{\frac{d^2u}{ds^2}\right\} \\
&= \left\{\frac{du}{ds}\right\}^2\left\{\frac{d^2G_N(u)}{du^2}\right\} + \left\{\frac{dG_N(u)}{du}\right\}\left\{\frac{d^2u}{ds^2}\right\},
\end{aligned}$$

$$\frac{d^2}{ds^2}M_Y(s)\Big|_{s=0} = E[Y^2] = \left[\left\{\frac{du}{ds}\right\}^2\left\{\frac{d^2G_N(u)}{du^2}\right\} + \left\{\frac{dG_N(u)}{du}\right\}\left\{\frac{d^2u}{ds^2}\right\}\right]_{s=0;\,u=1}$$
$$= \{-E[X]\}^2\{E[N^2]-E[N]\} + E[N]E[X^2]$$
$$= E[N^2]\{E[X]\}^2 + E[N]E[X^2] - E[N]\{E[X]\}^2.$$

The variance of Y is given by:

$$\begin{aligned}\sigma_Y^2 &= E[Y^2]-(E[Y])^2\\ &= E[N^2]\{E[X]\}^2 + E[N]E[X^2]-E[N]\{E[X]\}^2\\ &\quad -(E[N]E[X])^2\\ &= E[N]\left\{E[X^2]-\{E[X]\}^2\right\}+(E[X])^2\left\{E[N^2]-(E[N])^2\right\}\\ &= E[N]\sigma_X^2+(E[X])^2\sigma_N^2.\end{aligned}$$

If X is also a discrete random variable, then we obtain:

$$G_Y(z) = G_N(G_X(z)),$$

and the results for E[Y] and σ_Y^2 still hold.

1.7 SOME PROBABILITY DISTRIBUTIONS

Random variables with special probability distributions are encountered in different fields of science and engineering. In this section we describe some of these distributions, including their expected values, variances, and s-transforms (or z-transforms, as the case may be).

1.7.1 The Bernoulli Distribution

A Bernoulli trial is an experiment that results in two outcomes: *success* and *failure*. One example of a Bernoulli trial is the coin-tossing experiment, which results in heads or tails. In a Bernoulli trial we define the probability of success and probability of failure as follows:

$$P[\text{success}] = p \qquad 0 \le p \le 1$$
$$P[\text{failure}] = 1-p$$

Let us associate the events of the Bernoulli trial with a random variable X such that when the outcome of the trial is a success, we define $X = 1$, and when the outcome is a failure, we define $X = 0$. The random variable X is called a Bernoulli random variable, and its PMF is given by:

$$P_X(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \end{cases}$$

An alternative way to define the PMF of X is as follows:

$$p_X(x) = p^x(1-p)^{1-x} \quad x = 0, 1.$$

The CDF is given by:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

The expected value of X is given by:

$$E[X] = 0(1-p) + 1(p) = p.$$

Similarly, the second moment of X is given by:

$$E[X^2] = 0^2(1-p) + 1^2(p) = p.$$

Thus, the variance of X is given by:

$$\sigma_X^2 = E[X^2] - \{E[X]\}^2 = p - p^2 = p(1-p).$$

The z-transform of the PMF is given by:

$$G_X(z) = \sum_{x=0}^{\infty} z^x p_X(x) = \sum_{x=0}^{1} z^x p_X(x) = z^0(1-p) + z^1 p = 1 - p + zp.$$

1.7.2 The Binomial Distribution

Suppose we conduct n independent Bernoulli trials and we represent the number of successes in those n trials by the random variable $X(n)$. Then, $X(n)$ is defined as a binomial random variable with parameters (n, p). The PMF of a random variable, $X(n)$, with parameters (n, p) is given by:

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

The binomial coefficient, $\binom{n}{x}$, represents the number of ways of arranging x successes and $n - x$ failures.

The CDF, mean and variance of $X(n)$, and the z-transform of its PMF are given by:

$$\begin{aligned}
F_{X(n)}(x) &= P[X(n) \le x] = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}, \\
E[X(n)] &= np, \\
E[X^2(n)] &= n(n-1)p^2 + np, \\
\sigma^2_{X(n)} &= E[X^2(n)] - \{E[X(n)]\}^2 = np(1-p), \\
G_{X(n)}(z) &= (zp + 1 - p)^n.
\end{aligned}$$

1.7.3 The Geometric Distribution

The geometric random variable is used to describe the number of independent Bernoulli trials until the first success occurs. Let X be a random variable that denotes the number of Bernoulli trials until the first success. If the first success occurs on the xth trial, then we know that the first $x - 1$ trials resulted in failures. Thus, the PMF of a geometric random variable, X, is given by:

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

The CDF, mean, and variance of X and the z-transform of its PMF are given by:

$$\begin{aligned}
F_X(x) &= P[X \le x] = 1 - (1-p)^x, \\
E[X] &= 1/p, \\
E[X^2] &= \frac{2-p}{p^2}, \\
\sigma^2_X &= E[X^2] - \{E[X]\}^2 = \frac{1-p}{p^2}, \\
G_X(z) &= \frac{zp}{1 - z(1-p)}.
\end{aligned}$$

1.7.4 The Pascal Distribution

The Pascal random variable is an extension of the geometric random variable. A Pascal random variable of order k describes the number of trials until the kth success, which is why it is sometimes called the "kth-order interarrival time for a Bernoulli process." The Pascal distribution is also called the *negative binomial distribution*.

Let X_k be a kth-order Pascal random variable. Then its PMF is given by:

$$p_{X_k}(n) = \binom{n-1}{k-1} p^k (1-p)^{n-k} \quad k = 1, 2, \ldots; n = k, k+1, \ldots.$$

The CDF, mean, and variance of X_k and the z-transform of its PMF are given by:

$$
\begin{aligned}
F_{x_k}(x) &= P[X_k \le x] = \sum_{n=k}^{x} \binom{n-1}{k-1} p^k (1-p)^{n-k}, \\
E(X_k) &= k/p, \\
E[X_k^2] &= \frac{k^2 + k(1-p)}{p^2}, \\
\sigma_{X_k}^2 &= E[X_k^2] - \{E[X_k]\}^2 = \frac{k(1-p)}{p^2}, \\
G_{X_k}(z) &= \left[\frac{zp}{1-z(1-p)}\right]^k.
\end{aligned}
$$

1.7.5 The Poisson Distribution

A discrete random variable K is called a Poisson random variable with parameter λ, where $\lambda > 0$, if its PMF is given by:

$$p_K(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \ldots.$$

The CDF, mean, and variance of K and the z-transform of its PMF are given by:

$$
\begin{aligned}
F_K(k) &= P[K \le k] = \sum_{r=0}^{k} \frac{\lambda^r}{r!} e^{-\lambda}, \\
E[K] &= \lambda, \\
E[K^2] &= \lambda^2 + \lambda, \\
\sigma_K^2 &= E[K^2] - \{E[K]\}^2 = \lambda, \\
G_K(z) &= e^{\lambda(z-1)}.
\end{aligned}
$$

1.7.6 The Exponential Distribution

A continuous random variable X is defined to be an exponential random variable (or X has an exponential distribution) if for some parameter $\lambda > 0$ its PDF is given by:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

The CDF, mean, and variance of X and the s-transform of its PDF are given by:

$$
\begin{aligned}
F_X(x) &= P[X \le x] = 1 - e^{-\lambda x}, \\
E[X] &= 1/\lambda, \\
E[X^2] &= 2/\lambda^2, \\
\sigma_X^2 &= E[X^2] - \{E[X]\}^2 = 1/\lambda^2, \\
M_X(s) &= \frac{\lambda}{s+\lambda}.
\end{aligned}
$$

1.7.7 The Erlang Distribution

The Erlang distribution is a generalization of the exponential distribution. While the exponential random variable describes the time between adjacent events, the Erlang random variable describes the time interval between any event and the kth following event. A random variable is referred to as a kth-order Erlang (or Erlang-k) random variable with parameter λ if its PDF is given by:

$$
f_{X_k}(x) = \begin{cases} \dfrac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} & k = 1, 2, 3, \ldots; x \ge 0 \\ 0 & x < 0 \end{cases}
$$

The CDF, mean, and variance of X_k and the s-transform of its PDF are given by

$$
\begin{aligned}
F_{X_k}(x) &= P[X_k \le x] = 1 - \sum_{j=0}^{k-1} \frac{(\lambda x)^j e^{-\lambda x}}{j!}, \\
E[X_k] &= k/\lambda, \\
E[X_k^2] &= \frac{k(k+1)}{\lambda^2}, \\
\sigma_{X_k}^2 &= E[X_k^2] - \{E[X_k]\}^2 = \frac{k}{\lambda^2}, \\
M_{X_k}(s) &= \left[\frac{\lambda}{s+\lambda}\right]^k.
\end{aligned}
$$

1.7.8 The Uniform Distribution

A continuous random variable X is said to have a uniform distribution over the interval $[a, b]$ if its PDF is given by:

$$
f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}
$$

The CDF, mean, and variance of X and the s-transform of its PDF are given by:

$$F_X(x) = P[X \le x] = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x < b, \\ 1 & x \ge b \end{cases}$$

$$E[X] = \frac{b+a}{2},$$

$$E[X^2] = \frac{b^2 + ab + a^2}{3},$$

$$\sigma_X^2 = E[X^2] - \{E[X]\}^2 = \frac{(b-a)^2}{12},$$

$$M_X(s) = \frac{e^{-as} - e^{-bs}}{s(b-a)}$$

1.7.9 The Hyperexponential Distribution

The Erlang distribution belongs to a class of distributions that are said to have a *phase-type* distribution. This arises from the fact that the Erlang distribution is the sum of independent exponential distributions. Thus, an Erlang random variable can be thought of as the time to go through a sequence of phases or stages, each of which requires an exponentially distributed length of time. For example, since an Erlang-k random variable X_k is the sum of k exponentially distributed random variables X with mean $1/\mu$, and we can visualize X_k as the time it takes to complete a task that must go through k stages, where the time the task spends at each stage is X. Thus, we can represent the time to complete that task by the series of stages shown in Figure 1.1.

The hyperexponential distribution is another type of the phase-type distribution. The random variable H_k is used to model a process where an item can choose one of k branches. The probability that it chooses branch i is α_i, $i = 1, 2, \ldots, k$. The time it takes the item to traverse branch i is exponentially distributed with a mean of $1/\mu_i$. Thus, the PDF of H_k is given by:

$$f_{H_k}(x) = \sum_{i=1}^{k} \alpha_i \mu_i e^{-\mu_i x}, \quad x \ge 0$$

$$\sum_{i=1}^{k} \alpha_i = 1.$$

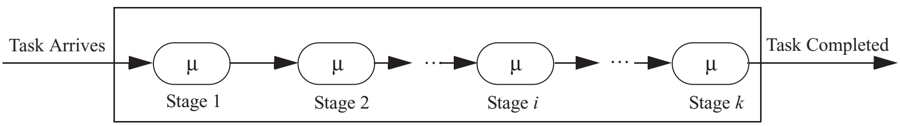

Figure 1.1 Graphical representation of the Erlang-k random variable.

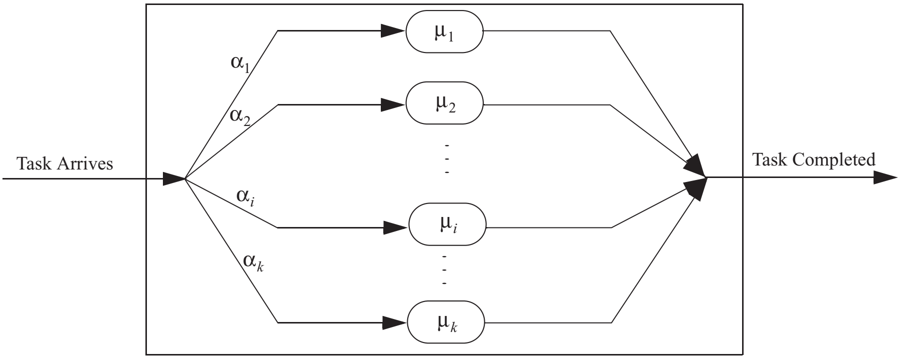

Figure 1.2 Graphical representation of H_k.

The random variable can be visualized as in Figure 1.2.

The mean, second moment, and s-transform of H_k are given by:

$$E[H_k] = \sum_{i=1}^{k} \frac{\alpha_i}{\mu_i},$$

$$E[H_k^2] = 2\sum_{i=1}^{k} \frac{\alpha_i}{\mu_i^2},$$

$$M_{H_k}(s) = \sum_{i=1}^{k} \frac{\alpha_i \mu_i}{s + \mu_i}.$$

1.7.10 The Coxian Distribution

The Coxian distribution is the third member of the phase-type distribution. A random variable C_k has a Coxian distribution of order k if it has to go through up to at most k stages, each of which has an exponential distribution. The random variable is popularly used to approximate general nonnegative distributions with exponential phases. The mean time spent at stage i is $1/\mu_i$, $i = 1, 2, \ldots, k$. A task arrives at stage 1; it may choose to receive some service at stage 1 with probability β_1 or leave the system with probability $\alpha_1 = 1 - \beta_1$. Given that it receives service at stage 1, the task may leave the system with probability α_2 or proceed to receive further service at stage 2 with probability $\beta_2 = 1 - \alpha_2$. This process continues until the task reaches stage k, where it finally leaves the system after service. The graphical representation of the process is shown in Figure 1.3.

The probability B_i of advancing to the ith stage to receive service is given by:

$$B_i = \prod_{j=1}^{i} \beta_j \quad i = 1, 2, \ldots, k.$$

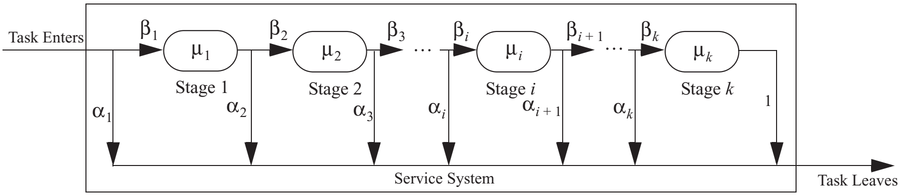

Figure 1.3 Graphical representation of C_k.

Thus, the probability L_i of leaving the system after the ith stage is given by:

$$L_i = \begin{cases} \alpha_{i+1} B_i & i = 1, 2, \ldots, k-1 \\ B_k & i = k \end{cases}$$

If we define the PDF of the service time X_i at stage i by $f_{X_i}(x)$, $x \geq 0$, then we note that C_k takes on the following values with the associated probabilities:

$$\begin{aligned} L_1 &= P[C_k = X_1], \\ L_2 &= P[C_k = X_1 + X_2], \\ L_3 &= P[C_k = X_1 + X_2 + X_3] = P\left[C_k = \sum_{i=1}^{3} X_i\right], \\ &\vdots \\ L_k &= P\left[C_k = \sum_{i=1}^{k} X_i\right]. \end{aligned}$$

Also, let $g_i(x)$ denote the PDF of the sum of random variables $X_1 + X_2 + \ldots + X_i$. Then, we know that $g_i(x)$ is the convolution of the PDFs of the X_i, that is,

$$g_i(x) = f_{X_1}(x) * f_{X_2}(x) * \ldots * f_{X_i}(x).$$

Therefore, the s-transform of $g_i(x)$ is:

$$M_{G_i}(s) = M_{X_1}(s) M_{X_2}(s) \ldots M_{X_i}(s) = \prod_{j=1}^{i} \frac{\mu_j}{s + \mu_j}.$$

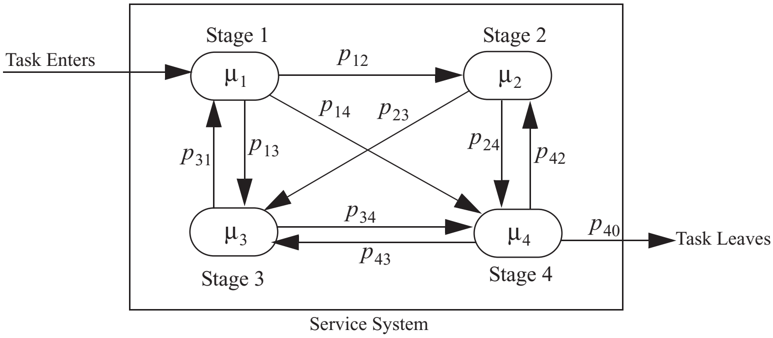

Figure 1.4 Graphical representation of phase-type distribution.

This means that the s-transform of C_k is given by:

$$M_{C_k}(s) = L_1 M_{G_1}(s) + L_2 M_{G_2}(s) + \ldots + L_k M_{G_k}(s) = \sum_{i=1}^{k} L_i M_{G_i}(s)$$
$$= \sum_{i=1}^{k} L_i \left\{ \prod_{j=1}^{\mathrm{i}} \frac{\mu_j}{s+\mu_j} \right\},$$

The mean and second moment of C_k are given by:

$$E[C_k] = \sum_{i=1}^{k} L_i \left\{ \prod_{j=1}^{\mathrm{i}} \frac{1}{\mu_j} \right\},$$
$$E[C_k^2] = 2\sum_{i=1}^{k} L_i \left\{ \sum_{j=1}^{i} \left[\frac{1}{\mu_j^2} + \sum_{\substack{l=1 \\ l \neq j}}^{i} \frac{1}{\mu_l \mu_j} \right] \right\}.$$

1.7.11 The General Phase-Type Distribution

The three types of phase-type distributions (Erlang, hyperexponential, and Coxian) are represented by feedforward networks of stages. A more general type of the phase-type distribution allows both feedforward and feedback relationships among the stages. This type is simply called the phase-type distribution. An example is illustrated in Figure 1.4.

This distribution is characterized by both the mean service time $1/\mu_i$ at stage i and a transition probability matrix that defines the probability p_{ij} that a task that has completed service at stage i goes next to stage j. The details of this particular type of distribution are very involved and will not be discussed here.

1.7.12 Normal Distribution

A continuous random variable X is defined to be a normal random variable with parameters μ_X and σ_X^2 if its PDF is given by:

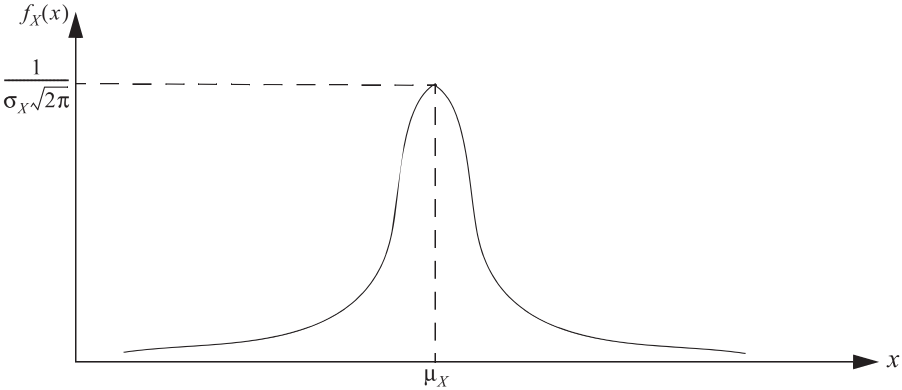

Figure 1.5 PDF of the normal random variable.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(x-\mu_X)^2/2\sigma_X^2} = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \quad -\infty < x < \infty.$$

The PDF is a bell-shaped curve that is symmetric about μ_X, which is the mean of X. The parameter σ_X^2 is the variance. Figure 1.5 illustrates the shape of the PDF.

The CDF of X is given by:

$$F_X(x) = P[X \le x] = \frac{1}{\sigma_X\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(u-\mu_X)^2/2\sigma_X^2} du.$$

The normal random variable X with parameters μ_X and σ_X^2 is usually designated $X = N(\mu_X, \sigma_X^2)$. The special case of zero mean and unit variance (i.e., $\mu_X = 0$ and $\sigma_X^2 = 1$) is designated $X = N(0, 1)$ and is called the *standard normal random variable*. Let $y = (u - \mu_X)/\sigma_X$. Then, $du = \sigma_X dy$ and the CDF of X becomes:

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu_X)/\sigma_X} e^{-y^2/2} dy.$$

Thus, with the above transformation, X becomes a standard normal random variable. The above integral cannot be evaluated in closed form. It is usually evaluated numerically through the function $\Phi(x)$, which is defined as follows:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

Thus, the CDF of X is given by

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu_X)/\sigma_X} e^{-y^2/2} dy = \Phi\left(\frac{x-\mu_X}{\sigma_X}\right).$$

The values of $\Phi(x)$ are sometimes given for nonnegative values of x. For negative values of x, $\Phi(x)$ can be obtained from the following relationship:

$$\Phi(-x) = 1 - \Phi(x).$$

Values of $\Phi(x)$ are given in standard books on probability, such as Ibe (2005).

1.8 LIMIT THEOREMS

In this section we discuss two fundamental theorems in probability. These are the law of large numbers, which is regarded as the first fundamental theorem, and the central limit theorem, which is regarded as the second fundamental theorem. We begin the discussion with the Markov and Chebyshev inequalities that enable us to prove these theorems.

1.8.1 Markov Inequality

The Markov inequality applies to random variables that take only nonnegative values. It can be stated as follows:

Proposition 1.1: If X is a random variable that takes only nonnegative values, then for any $a > 0$,

$$P[X \geq a] \leq \frac{E[X]}{a}.$$

Proof: We consider only the case when X is a continuous random variable. Thus,

$$\begin{aligned}
E[X] &= \int_0^{\infty} x f_X(x) dx = \int_0^{a} x f_X(x) dx + \int_a^{\infty} x f_X(x) dx \\
&\geq \int_a^{\infty} x f_X(x) dx \\
&\geq \int_a^{\infty} a f_X(x) dx \\
&= a\int_a^{\infty} f_X(x) dx \\
&= aP[X \geq a],
\end{aligned}$$

and the result follows.

1.8.2 Chebyshev Inequality

The Chebyshev inequality enables us to obtain bounds on probability when both the mean and variance of a random variable are known. The inequality can be stated as follows:

Proposition 1.2: Let X be a random variable with mean μ and variance σ^2. Then, for any $b > 0$,

$$P[|X-\mu| \geq b] \leq \frac{\sigma^2}{b^2}.$$

Proof: Since $(X - \mu)^2$ is a nonnegative random variable, we can invoke the Markov inequality, with $a = b^2$, to obtain:

$$P\left[(X-\mu)^2 \geq b^2\right] \leq \frac{E\left[(X-\mu)^2\right]}{b^2}.$$

Since $(X - \mu)^2 \geq b^2$ if and only if $|X - \mu| \geq b$, the preceding inequality is equivalent to:

$$P[|X-\mu| \geq b] \leq \frac{E\left[(X-\mu)^2\right]}{b^2} = \frac{\sigma^2}{b^2},$$

which completes the proof.

1.8.3 Law of Large Numbers

There are two laws of large numbers that deal with the limiting behavior of random sequences. One is called the "weak" law of large numbers and the other is called the "strong" law of large numbers. We will discuss only the weak law of large numbers.

Proposition 1.3: Let $X_1, X_2, \ldots, X_n$ be a sequence of mutually independent and identically distributed random variables, and let their mean be $E[X_k] = \mu < \infty$. Similarly, let their variance be $\sigma^2_{X_k} = \sigma^2 < \infty$. Let S_n denote the sum of the n random variables, that is,

$$S_n = X_1 + X_2 + \ldots + X_n.$$

Then the weak law of large numbers states that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left[\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right] \to 0.$$

Equivalently,

$$\lim_{n\to\infty} P\left[\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right] \to 1.$$

Proof: Since $X_1, X_2, \ldots, X_n$ are independent and have the same distribution, we have that:

$$\begin{aligned}
\operatorname{Var}(S_n) &= n\sigma^2, \\
\operatorname{Var}\left(\frac{S_n}{n}\right) &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}, \\
E\left[\frac{S_n}{n}\right] &= \frac{n\mu}{n} = \mu.
\end{aligned}$$

From Chebyshev inequality, for $\varepsilon > 0$, we have that:

$$P\left[\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right] \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Thus, for a fixed ε,

$$P\left[\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right] \to 0$$

as $n \to \infty$, which completes the proof.

1.8.4 The Central Limit Theorem

The central limit theorem provides an approximation to the behavior of sums of random variables. The theorem states that as the number of independent and identically distributed random variables with finite mean and finite variance increases, the distribution of their sum becomes increasingly normal regardless of the form of the distribution of the random variables. More formally, let $X_1, X_2, \ldots, X_n$ be a sequence of mutually independent and identically distributed random variables, each of which has a finite mean μ_X and a finite variance σ_X^2. Let S_n be defined as follows:

$$S_n = X_1 + X_2 + \ldots + X_n.$$

Now,

$$\begin{aligned}
E[S_n] &= n\mu_X, \\
\sigma_{S_n}^2 &= n\sigma_X^2.
\end{aligned}$$

Converting S_n to standard normal random variable (i.e., zero mean and variance = 1) we obtain:

$$Y_n = \frac{S_n - \overline{S}_n}{\sigma_{S_n}} = \frac{S_n - n\mu_X}{\sqrt{n\sigma_X^2}} = \frac{S_n - n\mu_X}{\sigma_X\sqrt{n}}.$$

The central limit theorem states that if $F_{Y_n}(y)$ is the CDF of Y_n, then:

$$\lim_{n\to\infty} F_{Y_n}(y) = \lim_{n\to\infty} P[Y_n \le y] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2} du = \Phi(y).$$

This means that $\lim_{n\to\infty} Y_n \sim N(0, 1)$. Thus, one of the important roles that the normal distribution plays in statistics is its usefulness as an approximation of other probability distribution functions.

An alternate statement of the theorem is that in the limit as n becomes very large,

$$\tilde{S}_n = \frac{S_n}{\sigma_{S_n}} = \frac{S_n}{\sigma_X \sqrt{n}} = \frac{X_1 + X_2 + \ldots + X_n}{\sigma_X \sqrt{n}}$$

is a normal random variable with unit variance.

1.9 PROBLEMS

1.1 A sequence of Bernoulli trials consists of choosing seven components at random from a batch of components. A selected component is classified as either defective or nondefective. A nondefective component is considered to be a success, while a defective component is considered to be a failure. If the probability that a selected component is nondefective is 0.8, what is the probability of exactly three successes?

1.2 The probability that a patient recovers from a rare blood disease is 0.3. If 15 people are known to have contracted this disease, find the following probabilities:

a. At least 10 survive.

b. From three to eight survive.

c. Exactly six survive.

1.3 A sequence of Bernoulli trials consists of choosing components at random from a batch of components. A selected component is classified as either defective or nondefective. A nondefective component is considered to be a success, while a defective component is considered to be a failure. If the probability that a selected component is nondefective is 0.8, determine the probabilities of the following events:

a. The first success occurs on the fifth trial.

b. The third success occurs on the eighth trial.

c. There are two successes by the fourth trial, there are four successes by the 10th trial, and there are 10 successes by the 18th trial.

1.4 A lady invites 12 people for dinner at her house. Unfortunately the dining table can only seat six people. Her plan is that if six or fewer guests come, then they will be seated at the table (i.e., they will have a sit-down dinner); otherwise, she will set up a buffet-style meal. The probability that each invited guest will come to dinner is 0.4, and each guest's decision is independent of other guests' decisions. Determine the following:

a. The probability that she has a sit-down dinner

b. The probability that she has a buffet-style dinner

c. The probability that there are at most three guests

1.5 A Girl Scout troop sells cookies from house to house. One of the parents of the girls figured out that the probability that they sell a set of packs of cookies at any house they visit is 0.4, where it is assumed that they sell exactly one set to each house that buys their cookies.

a. What is the probability that the first house where they make their first sale is the fifth house they visit?

b. Given that they visited 10 houses on a particular day, what is the probability that they sold exactly six sets of cookie packs?

c. What is the probability that on a particular day the third set of cookie packs is sold at the seventh house that the girls visit?

1.6 Students arrive for a lab experiment according to a Poisson process with a rate of 12 students per hour. However, the lab attendant opens the door to the lab when at least four students are waiting at the door. What is the probability that the waiting time of the first student to arrive exceeds 20 min? (By waiting time we mean the time that elapses from when a student arrives until the door is opened by the lab attendant.)

1.7 Cars arrive at a gas station according to a Poisson process at an average rate of 12 cars per hour. The station has only one attendant. If the attendant decides to take a 2-min coffee break when there were no cars at the station, what is the probability that one or more cars will be waiting when he comes back from the break, given that any car that arrives when he is on coffee break waits for him to get back?

1.8 An insurance company pays out claims on its life insurance policies in accordance with a Poisson process with an average rate of five claims per week. If the amount of money paid on each policy is uniformly distributed between \$2000 and \$10,000, what is the mean of the total amount of money that the company pays out in a 4-week period?

1.9 Three customers A, B, and C simultaneously arrive at a bank with two tellers on duty. The two tellers were idle when the three customers arrived, and A goes directly to one teller, B goes to the other teller, and C waits until either A or B leaves before she can begin receiving service. If the service times provided by the tellers are exponentially distributed with a mean of 4 min, what is the probability that customer A is still in the bank after the other two customers leave?

1.10 A five-motor machine can operate properly if at least three of the five motors are functioning. If the lifetime X of each motor has the PDF $f_X(x) = \lambda e^{-\lambda x}, x \geq 0, \lambda > 0$, and if the lifetimes of the motors are independent, what is the mean of the random variable Y, the time until the machine fails?