CHAPTER 1

REVIEW OF PROBABILITY THEORY

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, 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.

Elements of Random Walk and Diffusion Processes, First Edition. Oliver C. Ibe.

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,

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

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:

$$\begin{aligned}
&[X \le x] = \{w \mid X(w) \le x\} \\
&[X > x] = \{w \mid X(w) > x\} \\
&[a < X < x] = \{w \mid a < X(w) < b\}
\end{aligned}$$

These events have probabilities that are denoted by the following:

- $P[X \le x]$ 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 $[X \le x] = \{x \mid X(w) \le x\}$. The distribution function (or the cumulative distribution function (CDF)) of X is defined by

$$F_X(x) = P[X \le x], \quad -\infty < x < \infty \tag{1.1}$$

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. $0 \le F_X(x) \le 1$
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:

$$p_X(x) = P[X = x] \tag{1.2}$$

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:

$$F_X(x) = \sum_{k \leq x} p_X(k) \tag{1.3}$$

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,

$$P[X \in A] = \int_A f_X(x)dx \tag{1.4}$$

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

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{1.5}$$

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 0.
4. $P[X < a] = P[X \leq a] = \int_{-\infty}^{a} f_X(x)dx = F_X(a)$

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

$$E[X] = \begin{cases} \sum_i x_i p_X(x_i) & X \text{ discrete} \\ \int_{-\infty}^{\infty} x f_X(x)dx & X \text{ continuous} \end{cases} \tag{1.6}$$

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 $\overline{X}$ and by $\langle X \rangle$.

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

$$E[X^n] = \overline{X^n} = \begin{cases} \sum_i x_i^n p_X(x_i) & X \text{ discrete} \\ \int_{-\infty}^{\infty} x^n f_X(x)\,dx & X \text{ continuous} \end{cases} \tag{1.7}$$

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

$$E[(X - \overline{X})^n] = \overline{(X - \overline{X})^n} = \begin{cases} \sum_i (x_i - \overline{X})^n p_X(x_i) & X \text{ discrete} \\ \int_{-\infty}^{\infty} (x - \overline{X})^n f_X(x)dx & X \text{ continuous} \end{cases} \tag{1.8}$$

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,

$$\sigma_X^2 = E\left[\left(X-\overline{X}\right)^2\right] = \overline{\left(X-\overline{X}\right)^2} = \begin{cases} \sum_i \left(x_i-\overline{X}\right)^2 p_X(x_i) & X \text{ discrete} \\ \int_{-\infty}^{\infty} \left(x-\overline{X}\right)^2 f_X(x)dx & X \text{ continuous} \end{cases} \tag{1.9}$$

It can be shown that

$$\sigma_X^2 = E[X^2] - \{E[X]\}^2 \tag{1.10}$$

1.3 TRANSFORM METHODS

Different types of transforms are used in science and engineering. These include the z-transform, Laplace transform, and Fourier transform. One of the reasons for their popularity is that when they are introduced into the solution of many problems, the calculations become greatly simplified. For example, many solutions of equations that involve derivatives and integrals of functions are given as the convolution of two functions: $a(x) * b(x)$. As students of signal and systems know, the Fourier transform of a convolution is the product of the individual Fourier transforms. That is, if $F[g(x)]$ is the Fourier transform of the function $g(x)$, then

$$F[a(x) * b(x)] = A(w)\,B(w)$$

where $A(w)$ is the Fourier transform of $a(x)$ and $B(w)$ is the Fourier transform of $b(x)$. This means that the convolution operation can be replaced by the much simpler multiplication operation. In fact, sometimes transform methods are the only tools available for solving some types of problems.

We consider three types of transforms: the characteristic function of PDFs, the z-transform (or moment generating function) of PMFs, and the s-transform (or Laplace transform) of PDFs. The s-transform and the z-transform 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 Characteristic Function

Let $f_X(x)$ be the PDF of the continuous random variable X. The characteristic function of X is defined by

$$\Phi_X(w) = E[e^{jwX}] = \int_{-\infty}^{\infty} e^{jwx} f_X(x)dx \tag{1.11}$$

where $j = \sqrt{-1}$. We can obtain $f_X(x)$ from $\Phi_X(w)$ as follows:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_X(w) e^{-jwx} dw \tag{1.12}$$

If X is a discrete random variable with PMF $p_X(x)$, the characteristic function is given by

$$\Phi_X(w) = \sum_{x=-\infty}^{\infty} p_X(x) e^{jwx} \tag{1.13}$$

Note that $\Phi_X(0) = 1$, which is a test of whether a given function of w is a true characteristic function of the PDF or PMF of a random variable.

1.3.2 Moment-Generating Property of the Characteristic Function

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

$$\Phi_X(w) = \int_{-\infty}^{\infty} e^{jwx} f_X(x)\, dx$$

Taking the derivative of $\Phi_X(w)$, we obtain

$$\begin{aligned}
\frac{d}{dw}\Phi_X(w) &= \frac{d}{dw}\int_{-\infty}^{\infty} e^{jwx} f_X(x)dx = \int_{-\infty}^{\infty} \frac{d}{dw} e^{jwx} f_X(x)dx \\
&= \int_{-\infty}^{\infty} jxe^{jwx} f_X(x)dx \\
\frac{d}{dw}\Phi_X(w)\Big|_{w=0} &= \int_{-\infty}^{\infty} jxf_X(x)dx = jE[X] \\
\frac{d^2}{dw^2}\Phi_X(w) &= \frac{d}{dw}\int_{-\infty}^{\infty} jxe^{jwx} f_X(x)dx = -\int_{-\infty}^{\infty} x^2 e^{jwx} f_X(x)dx \\
\frac{d^2}{dw^2}\Phi_X(w)\Big|_{w=0} &= -\int_{-\infty}^{\infty} x^2 f_X(x)dx = -E[X^2]
\end{aligned}$$

In general,

$$\frac{d^n}{dw^n}\Phi_X(w)\Big|_{w=0} = j^n E[X^n] \tag{1.14}$$

1.3.3 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

$$M_X(s) = E[e^{-sX}] = \int_0^{\infty} e^{-sx} f_X(x)\, dx \tag{1.15}$$

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 $K=5$.

1.3.4 Moment-Generating Property of the *s*-Transform

As stated earlier, 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:

$$\begin{aligned}
\frac{d}{ds}M_X(s) &= \frac{d}{ds}\int_0^\infty e^{-sx} f_X(x)dx = \int_0^\infty \frac{d}{ds}e^{-sx} f_X(x)dx \\
&= -\int_0^\infty xe^{-sx} f_X(x)dx \\
\frac{d}{ds}M_X(s)\Big|_{s=0} &= -\int_0^\infty xf_X(x)dx = -E[X] \\
\frac{d^2}{ds^2}M_X(s) &= \frac{d}{ds}(-1)\int_0^\infty xe^{-sx} f_X(x)dx = \int_0^\infty x^2e^{-sx} f_X(x)dx \\
\frac{d^2}{ds^2}M_X(s)\Big|_{s=0} &= \int_0^\infty x^2 f_X(x)dx = E[X^2]
\end{aligned}$$

In general,

$$\frac{d^n}{ds^n}M_X(s)\Big|_{s=0} = (-1)^n E[X^n] \tag{1.16}$$

1.3.5 The *z*-Transform

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

$$G_X(z) = E[z^X] = \sum_{x=0}^\infty z^x p_X(x) \tag{1.17}$$

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,

$$G_X(z) = \sum_{x=0}^{\infty} z^x p_X(x) = p_X(0) + zp_X(1) + z^2 p_X(2) + z^3 p_X(3) + \cdots$$

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 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:

$$p_X(x) = \frac{1}{x!}\left[\frac{d^x}{dz^x} G_X(z)\right]_{z=0} \qquad x = 0, 1, 2, \ldots \tag{1.18}$$

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

1.3.6 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

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

Similarly,

$$\frac{d^2}{dz^2}G_X(z)=\frac{d}{dz}\sum_{x=1}^{\infty}xz^{x-1}p_X(x)=\sum_{x=1}^{\infty}x\frac{d}{dz}z^{x-1}p_X(x)$$
$$=\sum_{x=1}^{\infty}x(x-1)z^{x-2}p_X(x)$$
$$\frac{d^2}{dz^2}G_X(z)|_{z=1}=\sum_{x=1}^{\infty}x(x-1)p_X(x)=\sum_{x=0}^{\infty}x(x-1)p_X(x)$$
$$=\sum_{x=0}^{\infty}x^2p_X(x)-\sum_{x=0}^{\infty}xp_X(x)$$
$$=E[X^2]-E[X]$$
$$E[X^2]=\frac{d^2}{dz^2}G_X(z)|_{z=1}+\frac{d}{dz}G_X(z)|_{z=1}$$

Thus, the variance is obtained as follows:

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

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 $\mathrm{Cov}(X, Y)$ or σ_{XY}, is defined by

$$\begin{aligned}\mathrm{Cov}(X,Y)=\sigma_{XY}&=E[(X-\mu_X)(Y-\mu_Y)]\\&=E[XY-\mu_Y X-\mu_X Y+\mu_X\mu_Y]\\&=E[XY]-\mu_X\mu_Y\end{aligned}\tag{1.19}$$

If X and Y are independent, then $E[XY]=E[X]E[Y]=\mu_X\mu_Y$ and $\mathrm{Cov}(X, Y)=0$. However, the converse is not true; that is, if the covariance of X and Y is 0, it does not necessarily mean that X and Y are independent random variables. If the covariance of two random variables is 0, 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:

$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\sigma_X^2\sigma_Y^2}}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}\tag{1.20}$$

The correlation coefficient has the property that

$$-1\le\rho_{XY}\le 1\tag{1.21}$$

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 U can be used to model the reliability of systems with standby 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, U represents the time until the system fails, which is the sum of the lifetimes of both components.

Their CDF can be obtained as follows:

$$F_U(u) = P[U \leq u] = P[X+Y \leq u] = \iint_D f_{XY}(x, y)dx\,dy$$

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

Thus,

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

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

$$\begin{aligned} f_U(u) &= \frac{d}{du}F_U(u) = \frac{d}{du}\int_{-\infty}^{\infty} F_X(u-y)f_Y(y)dy = \int_{-\infty}^{\infty}\frac{d}{du}F_X(u-y)f_Y(y)dy \\ &= \int_{-\infty}^{\infty} f_X(u-y)f_Y(y)dy \end{aligned}$$

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

$$f_U(u) = f_X(u) * f_Y(u) \tag{1.22}$$

In general, if U 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

$$\begin{aligned} U &= X_1 + X_2 + \cdots + X_n \\ f_U(u) &= f_{X_1}(u) * f_{X_2}(u) * \cdots * f_{X_n}(u) \end{aligned} \tag{1.23}$$

Thus, the s-transform of the PDF of U is given by

$$M_U(s) = \prod_{i=1}^{n} M_{X_i}(s) \tag{1.24}$$

1.6 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 or characteristic functions, as the case may be).

1.6.1 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:

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

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

$$p_X(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \end{cases} \tag{1.25}$$

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

$$p_X(x) = p^x (1-p)^{1-x} \quad x = 0,1 \tag{1.26}$$

The CDF is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases} \tag{1.27}$$

The expected value of X is given by

$$E[X] = 0(1-p) + 1(p) = p \tag{1.28}$$

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

$$\sigma_X^2 = E[X^2] - \{E[X]\}^2 = p - p^2 = p(1-p) \tag{1.29}$$

The z-transform of the PMF is given by

$$G_X(z) = \sum_{x=0}^{\infty} z^x p_X(x) = z^0 p_X(0) + z^1 p_X(1) = 1 - p + zp \tag{1.30}$$

1.6.2 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

$$p_{X(n)}(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n \tag{1.31}$$

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

$$F_{X(n)}(x) = P[X(n) \le x] = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k} \tag{1.32a}$$

$$E[X(n)] = np \tag{1.32b}$$

$$\sigma_{X(n)}^2 = np(1-p) \tag{1.32c}$$

$$G_{X(n)}(z) = (1 - p + zp)^n \tag{1.32d}$$

1.6.3 The Geometric Distribution

The geometric random variable is used to describe the number of 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

$$p_X(x) = p(1-p)^{x-1} \qquad x = 1, 2, \ldots \tag{1.33}$$

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

$$F_X(x) = 1 - (1-p)^x \tag{1.34a}$$

$$E[X] = \frac{1}{p} \tag{1.34b}$$

$$\sigma_X^2 = \frac{1-p}{p} \tag{1.34c}$$

$$G_X(z) = \frac{zp}{1-z(1-p)} \tag{1.34d}$$

1.6.4 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

$$p_K(k) = \frac{\lambda^k}{k!}e^{-\lambda} \quad k = 0, 1, 2, \ldots \tag{1.35}$$

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

$$F_K(k) = \sum_{r=0}^{k} \frac{\lambda^r}{r!}e^{-\lambda} \tag{1.36a}$$

$$E[K] = \lambda \tag{1.36b}$$

$$\sigma_K^2 = \lambda \tag{1.36c}$$

$$G_K(z) = e^{-\lambda(1-z)} \tag{1.36d}$$

1.6.5 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

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0 \tag{1.37}$$

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

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0 \tag{1.38a}$$

$$E[X] = \frac{1}{\lambda} \tag{1.38b}$$

$$\sigma_X^2 = \frac{1}{\lambda^2} \tag{1.38c}$$

$$M_X(s) = \frac{\lambda}{s+\lambda} \tag{1.38d}$$

Assume that $X_1, X_2, \ldots, X_n$ is a set of independent and identically distributed exponential random variables with mean $E[X_i]=1/\lambda$. Let $X=X_1+X_2+\cdots+X_n$. Then X is defined as the nth-order *Erlang random variable*. One of the features of the exponential distribution is its *forgetfulness* property. Specifically,

$$P[X \le s+t \mid X > t] = P[X \le s] \Rightarrow f_{X|X>t}(x \mid X > t) = \lambda e^{-\lambda(x-t)}$$

1.6.6 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

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(x-\mu_X)^2/2\sigma_X^2} \qquad -\infty < x < \infty \tag{1.39}$$

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. 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 $X \sim N(\mu_X, \sigma_X^2)$. The special case of zero mean and unit variance (i.e., $\mu_X=0$ and $\sigma_X^2=1$) is designated $X \sim N(0, 1)$ 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 aforementioned transformation, X becomes a standard normal random variable. The aforementioned integral cannot be evaluated in closed form. It is usually evaluated numerically through the function $\Phi(x)$, which is defined as follows:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy \tag{1.40}$$

Thus, the CDF of X is given by

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu_X)/\sigma_X} e^{-y^2/2} dy = \Phi\left(\frac{x-\mu_x}{\sigma_X}\right) \tag{1.41}$$

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

$$\Phi(-x) = 1 - \Phi(x) \tag{1.42}$$

Values of $\Phi(x)$ are given in standard books on probability, such as Ibe (2005). The characteristic function of $f_X(x)$ is obtained as follows:

$$\Phi_X(w) = \int_{-\infty}^{\infty} e^{jwx} f_X(x)dx = \frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{-\infty}^{\infty} e^{jwx} e^{-(x-\mu_X)^2/2\sigma_X^2} dx$$

Let $u = (x - \mu_X)/\sigma_X$, which means that $x = u\sigma_X + \mu_X$ and $dx = \sigma_X du$. Thus,

$$\Phi_X(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} e^{jw(u\sigma_X + \mu_X)} dx = \frac{e^{jw\mu_X}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(u^2 - 2jwu\sigma_X)/2} dx$$

Now,

$$\frac{u^2 - 2jwu\sigma_X}{2} = \frac{u^2 - 2jwu\sigma_X + (jw\sigma_X)^2}{2} - \frac{(jw\sigma_X)^2}{2}$$
$$= \frac{(u - jw\sigma_X)^2}{2} + \frac{w^2\sigma_X^2}{2}$$

Thus,

$$\Phi_X(w) = \frac{e^{jw\mu_X}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(u^2 - 2jwu\sigma_X)/2} du = \frac{e^{jw\mu_X}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(u - jw\sigma_X)^2/2} e^{-w^2\sigma_X^2/2} du$$
$$= e^{(jw\mu_X - w^2\sigma_X^2/2)} \left\{ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(u - jw\sigma_X)^2/2} du \right\}$$

If we substitute $v = u - jw\sigma_X$, then the term in parentheses becomes

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-v^2/2} dv = 1$$

because the random variable $V \sim N(0, 1)$. Therefore,

$$\Phi_X(w) = e^{(jw\mu_X - w^2\sigma_X^2/2)} = \exp\left(jw\mu_X - \frac{w^2\sigma_X^2}{2} \right) \tag{1.43}$$

We state the following theorem whose proof can be found in any standard book on probability theory:

Theorem 1.1: Let $X_1, X_2, \ldots, X_n$ be independent normally distributed random variables with the mean and variance of X_i given by μ_i and σ_i^2, respectively. Then, the random variable

$$X = \sum_{i=1}^{n} X_i$$

has a normal distribution with mean and variance, respectively, given by

$$\mu = \sum_{i=1}^{n} \mu_i$$

$$\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$$

That is, if $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$, then $X \sim N(\mu, \sigma^2)$. For example, let X_1 and X_2 be independent and identically distributed standard random variables. We can compute $P[-1 \le X_1 + X_2 \le 3]$ by noting that according to the preceding theorem, $X = X_1 + X_2 \sim N(0, 2)$. Thus, if $F_X(x)$ is the CDF of X, we have that

$$\begin{aligned}
P[-1 \le X_1 + X_2 \le 3] = P[-1 \le X \le 3] &= F_X(3) - F_X(-1) \\
&= \Phi\left(\frac{3-0}{\sqrt{2}}\right) - \Phi\left(\frac{-1-0}{\sqrt{2}}\right) = \Phi\left(\frac{3\sqrt{2}}{2}\right) - \Phi\left(\frac{-\sqrt{2}}{2}\right) \\
&= \Phi(2.12) - \Phi(-0.71) = \Phi(2.12) - \{1 - \Phi(0.71)\} \\
&= \Phi(2.12) + \Phi(0.71) - 1 = 0.9830 + 0.7611 - 1 \\
&= 0.7441
\end{aligned}$$

1.7 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.

1.7.1 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 $a > 0$,

$$P[X \ge a] \le \frac{E[X]}{a} \tag{1.44}$$

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

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

and the result follows.

1.7.2 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 $b>0$,

$$P[|X-\mu| \geq b] \leq \frac{\sigma^2}{b^2} \tag{1.45}$$

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.

1.7.3 Laws 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,

$$S_n = X_1 + X_2 + \cdots + X_n$$

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

$$\lim_{n\to\infty} P\left[\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right] \to 0 \tag{1.46}$$

Equivalently,

$$\lim_{n\to\infty} P\left[\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right] \to 1 \tag{1.47}$$

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

$$\mathrm{Var}(S_n) = n\sigma^2$$
$$\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$
$$E\left[\frac{S_n}{n}\right] = \frac{n\mu}{n} = \mu$$

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.

1.7.4 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:

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

Now, $E[S_n] = n\mu_X$ and $\sigma_{S_n}^2 = n\sigma_X^2$. Converting S_n to standard normal random variable (i.e., zero mean and variance $=1$) we obtain

$$Y_n = \frac{S_n - E[S_n]}{\sigma_{S_n}} = \frac{S_n - n\mu_X}{\sqrt{n\sigma_X^2}} = \frac{S_n - n\mu_X}{\sigma_X\sqrt{n}} \tag{1.48}$$

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

$$\lim_{n\to\infty} F_{Y_n}(y) = \lim_{n\to\infty} P[Y_n \leq y] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2} du = \Phi(y) \tag{1.49}$$

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,

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{\sigma_X \sqrt{n}}$$

is a normal random variable with unit variance.

PROBLEMS

1.1 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:

1. The first success occurs on the fifth trial.
2. The third success occurs on the eighth trial.
3. There are 2 successes by the fourth trial, there are 4 successes by the tenth trial, and there are 10 successes by the eighteenth trial.

1.2 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:

1. The probability that she has a sit-down dinner
2. The probability that she has a buffet-style dinner
3. The probability that there are at most three guests

1.3 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.

1. What is the probability that the house where they make their first sale is the fifth house they visit?
2. Given that they visited 10 houses on a particular day, what is the probability that they sold exactly six sets of cookie packs?
3. 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?

1.4 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.)

1.5 Customers arrive at the neighborhood bookstore according to a Poisson process with an average rate of 10 customers per hour. Independently of other customers, each arriving customer buys a book with probability 1/8.

1. What is the probability that the bookstore sells no book during a particular hour?
2. What is the PDF of the time until the first book is sold?

1.6 Students arrive at the professor's office for extra help according to a Poisson process with an average rate of four students per hour. The professor does not start the tutorial until at least three students are available. Students who arrive while the tutorial is going on will have to wait for the next session.

1. Given that a tutorial has just ended and there are no students currently waiting for the professor, what is the mean time until another tutorial can start?
2. Given that one student was waiting when the tutorial ended, what is the probability that the next tutorial does not start within the first 2 h?

1.7 Consider 100 bulbs whose light times are independent and identically distributed exponential random variables with a mean of 5 h. Assume that these bulbs are used one at time and a failed bulb is immediately replaced by a new one. What is the probability that there is still a working bulb after 525 h?

1.8 Assume that X and Y are independent and identically distributed standard normal random variables. Compute the PDF of $V=1+X+Y$.

1.9 Assume that we roll a fair die one thousand times. Assume that the outcomes of the rolls are independent and let X be the number of fours that appear. Find the probability that $160 \leq X \leq 190$.