CHAPTER 19

OVERVIEW OF PROBABILITY AND STOCHASTIC PROCESSES

The comparatively late rise of the theory of probability shows how hard it is to grasp, and the many paradoxes show clearly that we, as humans, lack a well-grounded intuition in this matter.

In probability theory there is a great deal of art in setting up the model, in solving the problem, and in applying the results back to the real world actions that will follow.

—*The Art of Probability*, Richard Hamming

LEARNING OBJECTIVES

After studying this chapter, you should be able to:

- Understand the basic concepts of probability.
- Explain the concept of random variable.
- Understand some of the important basic concepts of stochastic processes.

Before setting out on our exploration of queueing analysis, we review background on probability and stochastic processes. The reader familiar with these topics can safely skip this chapter.

The chapter begins with an introduction to some elementary concepts from probability theory and random variables; this material is needed for Chapter 20, on queueing analysis. Following this, we look at stochastic processes, which are also relevant to queueing analysis.

19.1 PROBABILITY

We give here the barest outline of probability theory, but enough to support the rest of this chapter.

Definitions of Probability

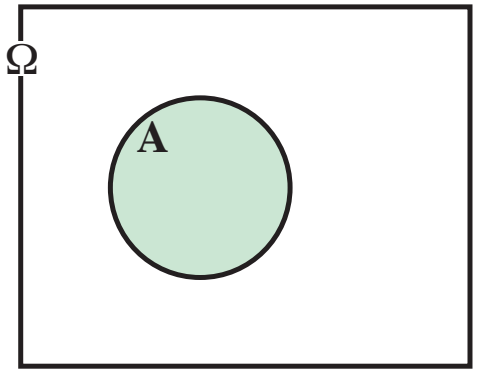
Probability is concerned with the assignment of numbers to events. The probability Pr[*A*] of an event *A* is a number between 0 and 1 that corresponds to the likelihood that the event *A* will occur. Generally, we talk of performing an experiment and obtaining an **outcome**. The **event** *A* is a particular outcome or set of outcomes, and a probability is assigned to that event.

It is difficult to get a firm grip on the concept of probability. Different applications of the theory present probability in different lights. In fact, there are a number of different definitions of probability. We highlight three here.

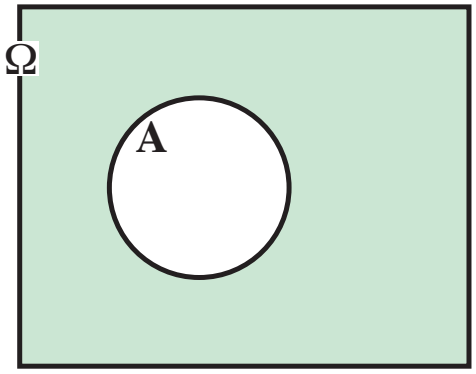
AXIOMATIC DEFINITION

A formal approach to probability is to state a number of axioms that define a probability measure and, from them, to derive laws of probability that can be used to perform useful calculations. The axioms are simply assertions that must be accepted. Once the axioms are accepted, it is possible to prove each of the laws.

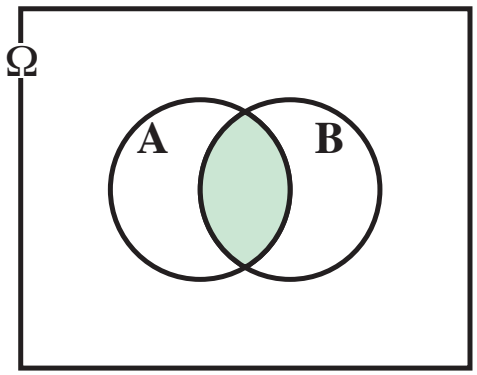
The axioms and laws make use of the following concepts from set theory. The **certain event** Ω is the event that occurs in every experiment; it consists of the universe, or **sample space**, of all possible outcomes. The **union** $A \cup B$ of two events *A* and *B* is the event that occurs when either *A* or *B* or both occur. The **intersection** $A \cap B$, also written *AB*, is the event that occurs when both events *A* and *B* occur. The events *A* and *B* are **mutually exclusive** if the occurrence of one of them excludes the occurrence of the other; that is, there is no outcome that is included in both *A* and *B*. The event $\overline{A}$, called the **complement** of *A*, is the event that occurs when *A* does not occur—that is, all outcomes in the sample space not included in *A*. These concepts are easily visualized with Venn diagrams, such as those shown in Figure 19.1. In each diagram, the shaded part corresponds to the expression below the diagram. Parts (c) and (d) correspond to cases in which *A* and *B*



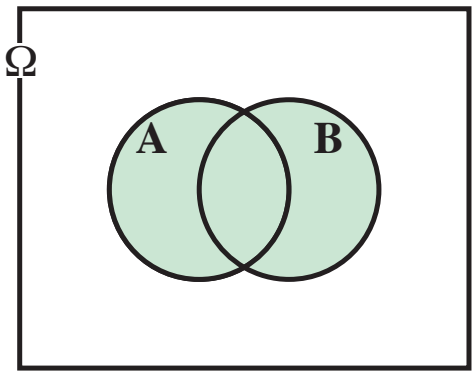
(a) A



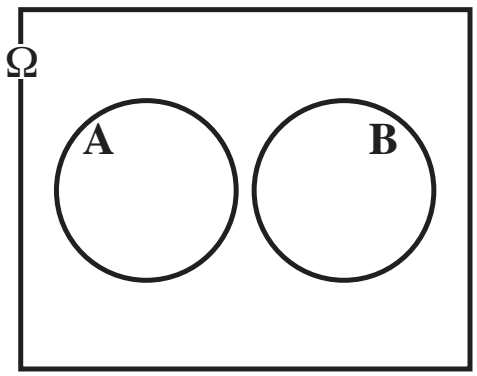
(b) NOT A
$\overline{A}$



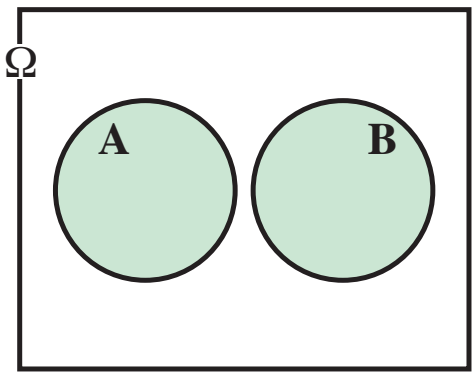
(c) A AND B
$A \cap B$



(d) A OR B
$A \cap B$



(e) A AND B
$A \cup B$



(f) A OR B
$A \cup B$

Figure 19.1 Venn Diagrams

are not mutually exclusive; that is, some outcomes are defined as part of both events *A* and *B*. Parts (e) and (f) correspond to cases in which *A* and *B* are mutually exclusive. Note that in these cases, the intersection of the two events is the empty set.

The common set of axioms used to define probability is as follows:

1. $0 \le \Pr[A] \le 1$ for each event *A*
2. $\Pr[\Omega] = 1$
3. $\Pr[A \cup B] = \Pr[A] + \Pr[B]$ if *A* and *B* are mutually exclusive

Axiom 3 can be extended to many events. For example, $\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B] + \Pr[C]$ if *A*, *B*, and *C* are mutually exclusive. Note that the axioms do not say anything about how probabilities are to be assigned to individual outcomes or events.

Based on these axioms, many laws can be derived. Here are some of the most important:

$\Pr[\overline{A}) = 1 - \Pr[A]$
$\Pr[A \cap B] = 0$ if *A* and *B* are mutually exclusive
$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$
$\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B] + \Pr[C] - \Pr[A \cap B] - \Pr[A \cap C] - \Pr[B \cap C] + \Pr[A \cap B \cap C]$

As an example, consider the throwing of a single die. This has six possible outcomes. The certain event is the event that occurs when any of the six die faces is on top. The union of the events {even} and {less than three} is the event {1 or 2 or 4 or 6}; the intersection of these events is the event {2}. The events {even} and {odd} are mutually exclusive. If we assume that each of the six outcomes is equally likely and assign a probability of 1/6 to each outcome, it is easy to see that the three axioms are satisfied. We can apply the laws of probability as follows:

Pr{even} = Pr{2} + Pr{4} + Pr{6} = 1/2

Pr{less than three} = Pr{1} + Pr{2} = 1/3

Pr[{even} ∪ {less than three}] = Pr{even} + Pr{less than three} – Pr{2}

= 1/2 + 1/3 – 1/6 = 2/3

Relative Frequency Definition

The relative frequency approach uses the following definition of probability. Perform an experiment a number of times; each time is called a **trial**. For each trial, observe whether the event *A* occurs. Then the probability Pr[*A*] of an event *A* is the limit:

$$\Pr[A] = \lim_{n \to \infty} \frac{n_A}{n}$$

where n is the number of trials and n_A is the number of occurrences of A.

For example, we could toss a coin many times. If the ratio of heads to total tosses hovers around 0.5 after a very large number of tosses, then we can assume that this is a fair coin, with equal probability of heads and tails.

Classical Definition

For the classical definition, let N be the number of possible outcomes, with the restriction that all outcomes are equally likely, and N_A the number of outcomes in which event A occurs. Then the probability of A is defined as:

$$\Pr[A] = \frac{N_A}{N}$$

For example, if we throw one die, then N is 6 and there are three outcomes that correspond to the event {even}; hence P{even} = 3/6 = 0.5. Here's a more complicated example: We roll two dice and want to determine the probability p that the sum is 7. You could consider the number of different sums that could be produced (2, 3, …, 12), which is 11, and conclude incorrectly that the probability is 1/11. We need to consider equally likely outcomes. For this purpose, we need to consider each combination of die faces, and we must distinguish between the first and second die. For example, the outcome (3, 4) must be counted separately from the outcome (4, 3). With this approach, there are 36 equally likely outcomes, and the favorable outcomes are the six pairs (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1). Thus, p = 6/36 = 1/6.

Conditional Probability and Independence

We often want to know a probability that is conditional on some event. The effect of the condition is to remove some of the outcomes from the sample space. For example, what is the probability of getting a sum of 8 on the roll of two dice, if we know that the face of at least one die is an even number? We can reason as follows. Because one die is even and the sum is even, the second die must show an even number. Thus, there are three equally likely successful outcomes: (2, 6), (4, 4) and (6, 2), out of a total set of possibilities of [36 – (number of events with both faces odd)] = 36 – 3 × 3 = 27. The resulting probability is 3/27 = 1/9.

Formally, the **conditional probability** of an event A assuming the event B has occurred, denoted by Pr[A|B], is defined as the ratio:

$$\Pr[A \mid B] = \frac{\Pr[AB]}{\Pr[B]}$$

where we assume Pr[B] is not zero.

In our example, A = {sum of 8} and B = {at least one die even}. The quantity Pr[AB] encompasses all of those outcomes in which the sum is 8 and at least one die is even. As we have seen, there are three such outcomes. Thus, Pr[AB] = 3/36 = 1/12. A moment's thought should convince you that Pr[B] = 3/4. We can now calculate:

$$\Pr[A \mid B] = \frac{1/12}{3/4} = \frac{1}{9}$$

This agrees with our previous reasoning.

Two events A and B are called **independent** if Pr[AB] = Pr[A]Pr[B]. It can easily be seen that if A and B are independent, Pr[A|B] = Pr[A] and Pr[B|A] = Pr[B].

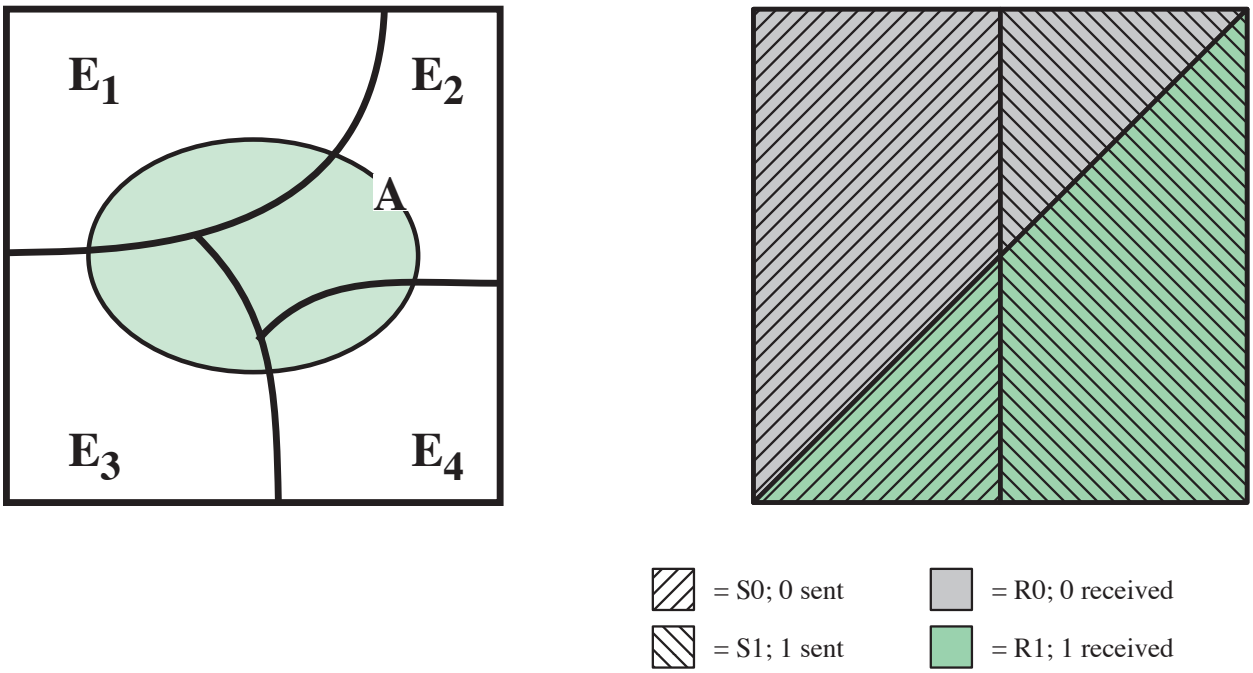
Bayes's Theorem

We close this section with one of the most important results from probability theory, known as Bayes's Theorem. First we need to state the total probability formula. Given a set of mutually exclusive events E_1, E_2, …, E_n, such that the union of these events covers all possible outcomes, and given an arbitrary event A, then it can be shown that

$$\Pr[A] = \sum_{i=1}^{n} \Pr[A \mid E_i] \Pr[E_i] \qquad \textbf{(19.1)}$$

Bayes's Theorem may be stated as follows:

$$\Pr[E_i \mid A] = \frac{\Pr[A \mid E_i] \mathrm{P}[E_i]}{\Pr[A]} = \frac{\Pr[A \mid E_i] \mathrm{P}[E_i]}{\sum_{j=1}^{n} \Pr[A \mid E_j] \Pr[E_j]}$$



(a) Diagram to illustrate concepts **(b) Example**

Figure 19.2 Illustration of Total Probability and Bayes' Theorem

Figure 19.2a illustrates the concepts of total probability and Bayes's Theorem.

Bayes's Theorem is used to calculate *posterior odds*, that is, the probability that something really is the case, given evidence in favor of it. For example, suppose we are transmitting a sequence of zeroes and ones over a noisy transmission line. Let S0 and S1 be the events that a 0 is sent at a given time and a 1 is sent, respectively, and R0 and R1 be the events that a 0 is received and a 1 is received. Suppose we know the probabilities of the source, namely Pr[S1] = p and Pr[S0] = 1 – p. Now the line is observed to determine how frequently an error occurs when a 1 is sent and when a 0 is sent, and the following probabilities are calculated: Pr[R0|S1] = p_a and Pr[R1|S0] = p_b. If a 0 is received, we can then calculate the conditional probability of an error, namely the conditional probability that a 1 was sent given that a 0 was received, using Bayes's Theorem:

$$\Pr[S1|R0] = \frac{\Pr[R0|S1]\Pr[S1]}{\Pr[R0|S1]\Pr[S1] + \Pr[R0|S0]\Pr[S0]} = \frac{p_a p}{p_a p + (1 - p_b)(1 - p)}$$

Figure 19.2b illustrates the preceding equation. In the figure, the sample space is represented by a unit square. Half of the square corresponds to SO and half to S1, so Pr[S0[= Pr[S1] = 0.5. Similarly, half of the square corresponds to RO and half to R1, so Pr[R0] = Pr[R1] = 0.5. Within the area representing S0, 1/4 of that area corresponds to R1, so Pr[R1/S0] = 0.25. Other conditional probabilities are similarly evident.

19.2 RANDOM VARIABLES

A **random variable** is a mapping from the set of all possible events in a sample space under consideration to the real numbers. That is, a random variable associates a real number with each event. This concept is sometimes expressed in terms of an experiment with many possible outcomes; a random variable assigns a value to each such outcome. Thus, the value of a random variable is a random quantity. We give the following formal definition. A random variable X is a function that assigns a number to every outcome in a sample space and satisfies the following conditions.

1. The set $\{X \leq x\}$ is an event for every x.
2. Pr[X = ∞] = Pr[X = −∞] = 0.

A random variable is **continuous** if it takes on an uncountably infinite number of distinct values. A random variable is **discrete** if it takes on a finite or countably infinite number of values.

Distribution and Density Functions

A continuous random variable X can be described by either its **distribution function** F(x) or **density function** f(x):

distribution function: $F(x) = \Pr[X \le x]$ $F(-\infty) = 0$; $F(\infty) = 1$

density function: $f(x) = \frac{d}{dx} F(x)$ $F(x) = \int_{-\infty}^{x} f(y)\,dy$ $\int_{-\infty}^{\infty} f(y)\,dy = 1$

For a discrete random variable, its probability distribution is characterized by

$$P_X(k) = \Pr[X = k] \quad \sum_{\text{all } k} P_X(k) = 1$$

We are often concerned with some characteristic of a random variable rather than the entire distribution, such as the following:

Mean value (also known as expected value or first moment):

$$\begin{cases} \mathrm{E}[X] = \mu_X = \int_{-\infty}^{\infty} x f(x)\,dx & \text{continuous case} \\ \mathrm{E}[X] = \mu_X = \sum_{\text{all } k} k \Pr[x = k] & \text{discrete case} \end{cases}$$

Second moment:

$$\begin{cases} \mathrm{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx & \text{continuous case} \\ \mathrm{E}[X^2] = \sum_{\text{all } k} k^2 \Pr[x = k] & \text{discrete case} \end{cases}$$

Variance $\mathrm{Var}[X] = \mathrm{E}\left[(X - \mu_X)^2\right] = \mathrm{E}[X^2] - \mu_X^2$

Standard deviation: $\sigma_X = \sqrt{\mathrm{Var}[X]}$

The variance and standard deviation are measures of the dispersion of values around the mean. A high variance means that the variable takes on

more values relatively farther from the mean than for a low variance. It is easy to show that for any constant a:

$$E[aX] = aE[X]; \qquad Var[aX] = a^2 Var[X]$$

The mean is known as a first-order statistic; the second moment and variance are second-order statistics. Higher-order statistics can also be derived from the probability density function.

Important Distributions

Several distributions that play an important role in queueing analysis are described next.

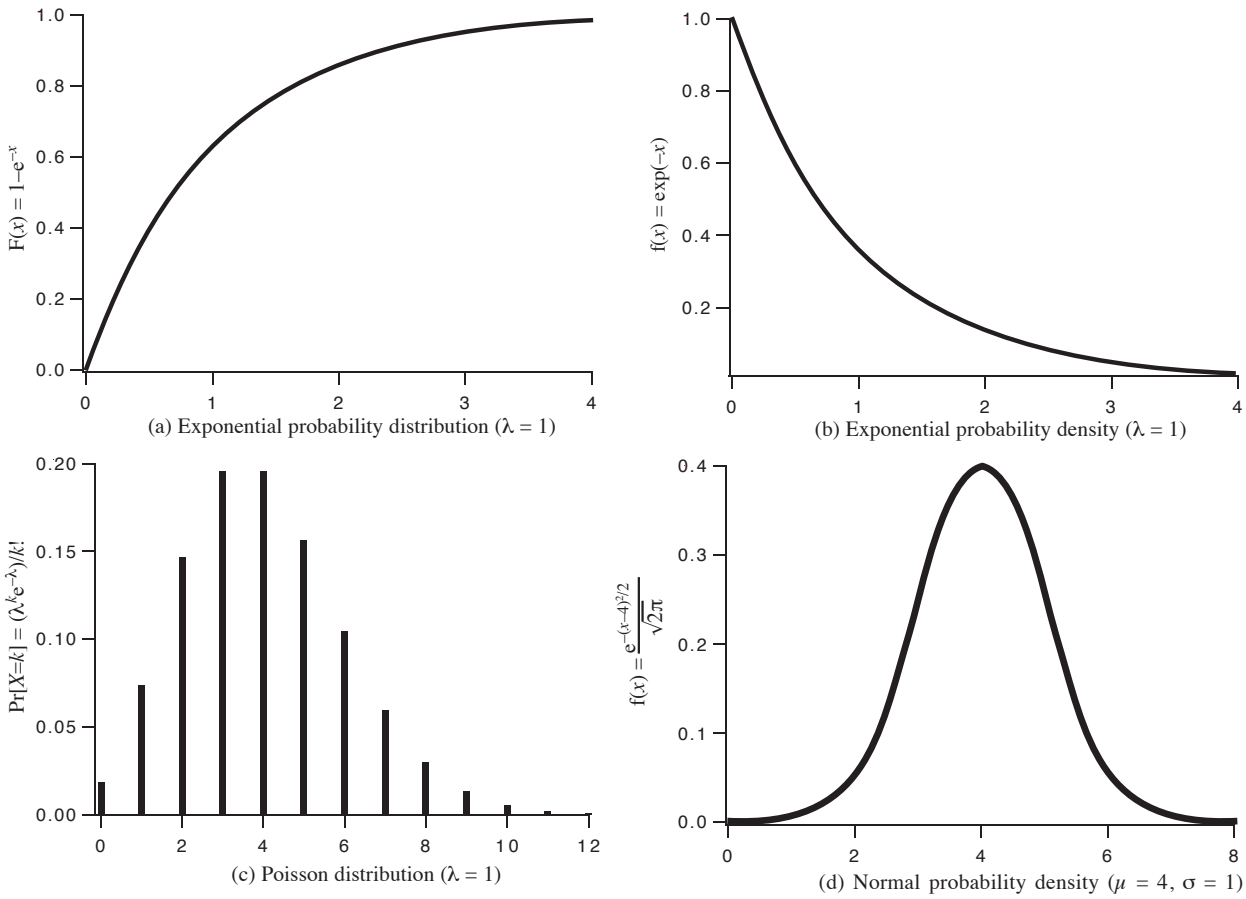


Figure 19.3 Some Probability Functions

EXPONENTIAL DISTRIBUTION

The exponential distribution with parameter $\lambda > 0$ is given by (Figures 19.3a and 19.3b) and has the following distribution and density functions:

$$F(x) = 1 - e^{-\lambda x} \qquad f(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

The exponential distribution has the interesting property that its mean is equal to its standard deviation:

$$\mathrm{E}[X] = \sigma_X = \frac{1}{\lambda}$$

When used to refer to a time interval, such as a service time, this distribution is sometimes referred to as a random distribution. This is because, for a time interval that has already begun, each time at which the interval may finish is equally likely.

This distribution is important in queueing theory because we can often assume that the service time of a server in a queueing system is exponential. In the case of telephone traffic, the service time is the time for which a subscriber engages the equipment of interest. In a packet-switching network, the service time is the transmission time and is therefore proportional to the packet length. It is difficult to give a sound theoretical reason why service times should be exponential, but in many cases they are very nearly exponential. This is good news because it simplifies the queueing analysis immensely.

POISSON DISTRIBUTION

Another important distribution is the Poisson distribution (Figure 19.3c), with parameter $\lambda > 0$, which takes on values at the points 0, 1,…:

$$\Pr[X = k] = \frac{\lambda^k}{k!} e^{-\lambda} \qquad k = 0, 1, 2\ldots$$

$$E[X] = \mathrm{Var}[X] = \lambda$$

If $\lambda < 1$, then $\Pr[X = k]$ is maximum for $k = 0$. If $\lambda > 1$ but not an integer, then $\Pr[X = k]$ is maximum for the largest integer smaller than λ; if λ is a positive integer, then there are two maxima at $k = \lambda$ and $k = \lambda - 1$.

The Poisson distribution is also important in queueing analysis because we must assume a Poisson arrival pattern to be able to develop the queueing equations (discussed in Chapter 20). Fortunately, the assumption of Poisson arrivals is usually valid.

The way in which the Poisson distribution can be applied to arrival rate is as follows. If items arrive at a queue according to a Poisson process, this may be expressed as:

$$\Pr[k \text{ items arrive in time interval } T] = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$$

E[number of items to arrive in time interval T] = λT

Mean arrival rate, in items per second = λ

Arrivals occurring according to a Poisson process are often referred to as random arrivals. This is because the probability of arrival of an item in a small interval is proportional to the length of the interval and is independent of the amount of elapsed time since the arrival of the last item. That is, when items are arriving according to a Poisson process, an item is as likely to arrive at one instant as any other, regardless of the instants at which the other customers arrive.

Another interesting property of the Poisson process is its relationship to the exponential distribution. If we look at the times between arrivals of items T_a (called the interarrival times), then we find that this quantity obeys the exponential distribution:

$$\Pr[T_a < t] = 1 - e^{-\lambda t}$$

$$E[T_a] = \frac{1}{\lambda}$$

Thus, the mean interarrival time is the reciprocal of the arrival rate, as we would expect.

NORMAL DISTRIBUTION

The normal distribution with parameters $\mu > 0$ and σ has the following density function (Figure 19.3d) and distribution function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(y-\mu)^2/2\sigma^2}\, dy$$

with

$E[X] = \mu$

$Var[X] = \sigma^2$

An important result is the central limit theorem, which states that the distribution of the average of a large number of independent random variables will be approximately normal, almost regardless of their individual distributions. One key requirement is finite mean and variance. The central limit theorem plays a key role in statistics.

Multiple Random Variables

With two or more random variables, we are often concerned whether variations in one are reflected in the other. This subsection defines some important measures of dependence.

In general, the statistical characterization of multiple random variables requires a definition of their joint probability density function or joint probability distribution function:

Distribution: $(F\, x_1, x_2, \ldots x_n) = \Pr[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n]$

Density: $f(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F(x_1, x_2, \ldots, x_n)$

Discrete distribution: $P(x_1, x_2, \ldots x_n) = \Pr[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]$

For any two random variables X and Y, we have

$$E[X + Y] = E[X] + E[Y]$$

Two continuous random variables X and Y are called (statistically) **independent** if $F(x, y) = F(x)F(y)$, and therefore $f(x, y) = f(x)f(y)$. If the random variables X and Y are discrete, then they are independent if $P(x, y) = P(x)P(y)$.

For independent random variables, the following relationships hold:

$$E[X\,Y] = E[X] \times E[Y]$$
$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

The **covariance** of two random variables X and Y is defined as follows:

$$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y]$$

If the variances of X and Y are finite, then their covariance is finite but may be positive, negative, or zero.

For finite variances of X and Y, the **correlation coefficient** of X and Y is defined as:

$$r(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \qquad \textbf{(19.2)}$$

We can think of this as a measure of the linear dependence between X and Y, normalized to be relative to the amount of variability in X and Y. The following relationship holds:

$$-1 \leq r(X, Y) \leq 1$$

It is said that X and Y are **positively correlated** if $r(X, Y) > 0$, that X and Y are **negatively correlated** if $r(X, Y) < 0$, and that X and Y are **uncorrelated** if $r(X, Y) = \text{Cov}(X, Y) = 0$. If X and Y are independent random variables, then they are uncorrelated and $r(X, Y) = 0$. However, it is possible for X and Y to be uncorrelated but not independent (see Problem 19.12).

The correlation coefficient provides a measure of the extent to which two random variables are linearly related. If the joint distribution of X and Y is relatively concentrated around a straight line in the xy-plane that has a positive slope, then $r(X, Y)$ will typically be close to 1. This indicates that a movement in X will be matched by a movement of relatively similar magnitude and direction in Y. If the joint distribution of X and Y is relatively concentrated around a straight line that has a negative slope, then $r(X, Y)$ will typically be close to −1.

The following relationship is easily demonstrated:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

If X and Y have the same variance σ^2, then the preceding can be rewritten as:

$$\text{Var}(X + Y) = 2\sigma^2(1 + r(X, Y))$$

If X and Y are uncorrelated [$r(X, Y) = 0$], then $\text{Var}(X + Y) = 2\sigma^2$. These results easily generalize to more than two variables: Consider a set of random variables $X_1, \ldots, X_N$, such that each has the same variance σ^2. Then

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = \sigma^2\left(N + 2\sum_i \sum_{j<i} r(i,j)\right)$$

where $r(i, j)$ is shorthand for $r(X_i, X_j)$. Using the relationship $\mathrm{Var}(X/N) = \mathrm{Var}(X)/N^2$, we can develop an equation for the variance of the sample mean of a set of random variables:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$\mathrm{Var}\left(\overline{X}\right) = \frac{\sigma^2}{N}\left(1 + \sum_i \sum_{j<i} r(i,j)\right)$$

If the X_i are mutually independent, then we have $\mathrm{Var}\left(\overline{X}\right) = \dfrac{\sigma^2}{N}$.

19.3 ELEMENTARY CONCEPTS OF STOCHASTIC PROCESSES

A **stochastic process**, also called a **random process**, is a family of random variables $\{\mathbf{x}(t), t \in T\}$ indexed by a parameter t over some index set T. Typically, the index set is interpreted as the time dimension, and $\mathbf{x}(t)$ is a function of time. Another way to say this is that a stochastic process is a random variable that is a function of time. A **continuous-time stochastic process** is one in which t varies continuously, typically over the nonnegative real line $\{\mathbf{x}(t), 0 \le t < \infty\}$, although sometimes over the entire real line; whereas a **discrete-time stochastic process** is one in which t takes on discrete values, typically the positive integers $\{\mathbf{x}(t), t = 1, 2, \ldots\}$, although in some cases the range is the integers from $-\infty$ to $+\infty$.

Recall that a random variable is defined as a function that maps the outcome of an experiment into a given value. With that in mind, the expression $\mathbf{x}(t)$ can be interpreted in several ways:

1. A family of time functions (t variable; all possible outcomes)
2. A single time function (t variable; one outcome)
3. A random variable (t fixed; all possible outcomes)
4. A single number (t fixed; one outcome)

The specific interpretation of $\mathbf{x}(t)$ is usually clear from the context.

A word about terminology. A **continuous-value stochastic process** is one in which the random variable $\mathbf{x}(t)$ with t fixed (case 3) takes on continuous values, whereas a **discrete-value stochastic process** is one in which the random variable at any time t takes on a finite or countably infinite number of values. A continuous-time stochastic process may be either continuous value or discrete value, and a discrete-time stochastic process may be either continuous value or discrete value.

As with any random variable, $\mathbf{x}(t)$ for a fixed value of t can be characterized by a probability distribution and a probability density. For continuous-value stochastic processes, these functions take the following form:

distribution function: $(x;\ t)\ F = \Pr[\mathbf{x}(t) \le x]\quad F(-\infty;\ t) = 0;\quad F(\infty;\ t) = 1$

density function: $f(x;t) = \dfrac{\partial}{\partial x}\mathrm{F}(x;t) \qquad F(x;t) = \int_{-\infty}^{x} f(y;t)\,dy \qquad \int_{-\infty}^{\infty} f(y;t)\,dy = 1$

For discrete-value stochastic processes:

$$P_{\mathbf{x}(t)}(k) = \Pr[\mathbf{x}(t) = k] \qquad \sum_{\text{all } k} P_{\mathbf{x}(t)}(k) = 1$$

A full statistical characterization of a stochastic process must take into account the time variable. Using the first interpretation in the preceding list, a stochastic process $\mathbf{x}(t)$ comprises an infinite number of random variables, one for each t. To specify fully the statistics of the process, we would need to specify the joint probability density function of the variables $\mathbf{x}(t_1)$, $\mathbf{x}(t_2)$, …, $\mathbf{x}(t_n)$ for all values of n ($1 \le n < \infty$) and all possible sampling times (t_1, t_2, …, t_n). For our purposes, we need not pursue this topic.

First- and Second-Order Statistics

The mean and variance of a stochastic process are defined in the usual way;

$$\mathrm{E}\left[\mathbf{x}(t)\right] = \mu(t) = \int_{-\infty}^{\infty} x f(x;t)\,dx \qquad \text{continuous-value case}$$

$$\mathrm{E}\left[\mathbf{x}(t)\right] = \mu(t) = \sum_{\text{all } k} k \Pr\left[x(t) = k\right] \qquad \text{discrete-value case}$$

$$\mathrm{E}\left[\mathbf{x}^2(t)\right] = \int_{-\infty}^{\infty} x^2 f(x;t)\,dx \qquad \text{continuous-value case}$$

$$\mathrm{E}\left[\mathbf{x}^2(t)\right] = \sum_{\text{all } k} k^2 \Pr\left[x(t) = k\right] \qquad \text{discrete-value case}$$

$$Var\left[\mathbf{x}(t)\right] = \sigma^2_{\mathbf{x}(t)} = \mathrm{E}\left[\left(\mathbf{x}(t) - \mu(t)\right)^2\right] = \mathrm{E}\left[\mathbf{x}^2(t)\right] - \mu^2(t)$$

Note that, in general, the mean and variance of a stochastic process are functions of time. An important concept for our discussion is the **autocorrelation function** $R(t_1, t_2)$, which is the joint moment of the random variables $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$:

$$R(t_1, t_2) = \mathrm{E}[\mathbf{x}(t_1)\mathbf{x}(t_2)]$$

As with the correlation function for two random variables introduced earlier, the autocorrelation is a measure of the relationship between the two time instances of a stochastic process. A related quantity is the **autocovariance**:

$$C(t_1, t_2) = \mathrm{E}[(\mathbf{x}(t_1) - \mu(t_1))(\mathbf{x}(t_2) - \mu(t_2))] = R(t_1, t_2) - \mu(t_1)\,\mu(t_2) \quad \textbf{(19.3)}$$

Note that the variance of $\mathbf{x}(t)$ is given by:

$$\mathrm{Var}[\mathbf{x}(t)] = C(t, t) = R(t, t) - \mu^2(t)$$

Finally, the **correlation coefficient** (see Equation 19.2) of $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$ is called the normalized autocorrelation function of the stochastic process and can be expressed as:

$$\rho(t_1,t_2) = \frac{\mathrm{E}\left[\left(x(t_1)-\mu(t_1)\right)\left(x(t_2)-\mu(t_2)\right)\right]}{\sigma_1\sigma_2} = \frac{C(t_1,t_2)}{\sigma_1\sigma_2} \quad \textbf{(19.4)}$$

Unfortunately, some texts and some of the literature refer to $\rho(t_1, t_2)$ as the autocorrelation function, so the reader must beware.

Stationary Stochastic Processes

In general terms, a **stationary stochastic process** is one in which the probability characteristics of the process do not vary as a function of time. There are several different precise definitions of this concept, but the one of most interest here is the concept of **wide sense stationary**. A process is

stationary in the wide sense (or weakly stationary) if its expected value is a constant and its autocorrelation function depends only on the time difference:

$$\mathrm{E}[\mathbf{x}(t)] = \mu$$
$$R(t, t+\tau) = R(t+\tau, t) = R(\tau) = R(-\tau) \quad \text{for all } t$$

From these equalities, the following can be derived:

$$\mathrm{Var}[\mathbf{x}(t)] = R(t, t) - \mu^2(t) = R(0) - \mu^2$$
$$C(t, t+\tau) = R(t, t+\tau) - \mu(t)\,\mu(t+\tau) = R(\tau) - \mu^2 = C(\tau)$$

An important characteristic of $R(\tau)$ is that it measures the degree of dependence of one time instant of a stochastic process on other time instants. If $R(\tau)$ goes to zero exponentially fast as τ becomes large, then there is little dependence of one instant of a stochastic process on instants far removed in time. Such a process is called a **short memory process**, whereas if $R(\tau)$ remains substantial for large values of τ(decays to zero at a slower than exponential rate), the stochastic process is a **long memory process**.

Spectral Density

The **power spectrum**, or **spectral density**, of a stationary random process is the Fourier transform of its autocorrelation function:

$$S(w) = \int_{-\infty}^{\infty} R(\tau) e^{-\mathrm{j}w\tau} d\tau$$

where w is the frequency in radians ($w = 2\pi f$) and $\mathrm{j} = \sqrt{-1}$.

For a deterministic time function, the spectral density gives the distribution frequency of the power of the signal. For a stochastic process, $S(w)$ is the average density of power in the frequency components of $\mathbf{x}(t)$ in the neighborhood of w. Recall that one interpretation of $\mathbf{x}(t)$ is that of a single time function (t variable; one outcome). For that interpretation, the time function, as with any time function, is made up of a summation of frequency components, and its spectral density gives the relative power contributed by each component. If we view $\mathbf{x}(t)$ as a family of time functions (t variable; all possible outcomes), then the spectral density gives the average power in each frequency component, averaged over all possible time functions $\mathbf{x}(t)$.

The Fourier inversion formula gives the time function in terms of its Fourier transform:

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\mathrm{w}) e^{\mathrm{j}w\tau} dw$$

With $\tau = 0$, the preceding yields:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} S(\mathrm{w}) dw = R(0) = \mathrm{E}\left[\left| \mathbf{x}(t) \right|^2 \right]$$

Thus, the total area under $S(\mathrm{w})/2\pi$ equals the average power of the process $\mathbf{x}(t)$. Also note:

$$S(0) = \int_{-\infty}^{\infty} R(\tau) d\tau$$

$S(0)$ represents the direct-current (dc) component of the power spectrum and corresponds to the integral of the autocorrelation function. This component will be finite only if $R(\tau)$ decays as $\tau \to \infty$ sufficiently rapidly for the integral of $R(\tau)$ to be finite.

We can also express the power spectrum for a stochastic process that is defined at discrete points in time (discrete-time stochastic process). In this case, we have:

$$S(w) = \sum_{k=-\infty}^{\infty} R(k) e^{-jkw} \quad S(0) = \sum_{k=-\infty}^{\infty} R(k)$$

Again, $S(0)$ represents the dc component of the power spectrum and corresponds to the infinite sum of the autocorrelation function. This component will be finite only if $R(\tau)$ decays as $\tau \to \infty$ sufficiently rapidly for the summation to be finite.

Table 19.1 shows some interesting correspondences between the autocorrelation function and the power spectral density.

Table 19.1 Autocorrelation Functions and Spectral Densities

Stationary Random Process	Autocorrelation Function	Power Spectral Density
$X(t)$	$R_X(\tau)$	$S_X(w)$
$aX(t)$	$a^2 R_X(\tau)$	$a^2 S_X(w)$
$X'(t)$	$-d^2 R_X(\tau)/d\tau^2$	$w^2 S_X(w)$
$X^{(n)}(t)$	$(-1)^n d^{2n} R_X(\tau)/d\tau^{2n}$	$w^{2n} S_X(w)$
$X(t)\exp(jw_0 t)$	$\exp(jw_0\tau) R_X(\tau)$	$S_X(w - w_0)$

Independent Increments

A continuous-time stochastic process $\{\mathbf{x}(t), 0 \le t < \infty\}$ is said to have independent increments if $\mathbf{x}(0) = 0$ and, for all choices of indexes $t_0 < t_1 < \ldots < t_n$, the n random variables

$$\mathbf{x}(t_1) - \mathbf{x}(t_0), \mathbf{x}(t_2) - \mathbf{x}(t_1), \ldots, \mathbf{x}(t_n) - \mathbf{x}(t_{n-1})$$

are independent. Thus, the amount of "movement'' in a stochastic process in one time interval is independent of the movement in any other nonoverlapping time interval. The process is said to have stationary independent increments if, in addition, $\mathbf{x}(t_2 + h) - \mathbf{x}(t_1 + h)$ has the same distribution as $\mathbf{x}(t_2) - \mathbf{x}(t_1)$ for all choices of $t_2 > t_1$ and every $h > 0$.

Two properties of processes with stationary independent increments are noteworthy. If $\mathbf{x}(t)$ has stationary independent increments and $E[\mathbf{x}(t)] = \mu(t)$ is a continuous function of time, then $\mu(t) = a + bt$, where a and b are constants. Also, if $\mathrm{Var}[\mathbf{x}(t) - \mathbf{x}(0)]$ is a continuous function of time, then for all s, $\mathrm{Var}[\mathbf{x}(s+ t) - \mathbf{x}(s)] = \sigma^2 t$, where σ^2 is a constant.

Two processes that play a central role in the theory of stochastic processes, the Brownian motion process and the Poisson process, have independent increments. A brief introduction to both follows.

Brownian Motion Process

Brownian motion is the random movement of microscopic particles suspended in a liquid or gas, caused by collisions with molecules of the surrounding medium. This physical phenomenon is the basis for the definition of the Brownian motion stochastic process, also known as the Wiener process and the Wiener-Levy process.

Let us consider the function $B(t)$ for a particle in Brownian motion as denoting the displacement from a starting point in one dimension after time t. Consider the net movement of the particle in a time interval (s, t), which is long compared to the time between impacts. The quantity $B(t) - B(s)$ can be viewed as the sum of a large number of small displacements. By the central limit theorem, we can assume that this quantity has a normal probability distribution.

If we assume that the medium is in equilibrium, it is reasonable to assume that the net displacement depends only on the length of the time interval and not on the time at which the interval begins. That is, the probability distribution of $B(t) - B(s)$ should be the same as $B(t + h) - B(s + h)$ for any $h > 0$. Finally, if the motion of the particle is due entirely to frequent random collisions, then the net displacements in nonoverlapping time intervals should be independent, and therefore $B(t)$ has independent increments.

With the foregoing reasoning in mind we define a Brownian motion process $B(t)$ as one that satisfies the following conditions:

1. $\{B(t), 0 \le t < \infty\}$ has stationary independent increments.
2. For every $t > 0$, the random variable $B(t)$ has a normal distribution.
3. For all $t > 0$, $E[B(t)] = 0$.
4. $B(0) = 0$.

The probability density of a Brownian motion process has the form:

$$f_B(x,t) = \frac{1}{\sigma\sqrt{2\pi t}} e^{-x^2/2\sigma^2 t}$$

From this we have:

Var[$B(t)$] = t; Var[$B(t) - B(s)$] = $|t - s|$

Another important quantity is the autocorrelation of $B(t)$, expressed as $R_B(t_1, t_2)$. We derive this quantity in the following way. First, observe that for $t_4 > t_3 > t_2 > t_1$:

$$\begin{aligned}\mathrm{E}\left[\left(B(t_4)-B(t_3)\right)\left(B(t_2)-B(t_1)\right)\right] &= \mathrm{E}\left[B(t_4)-B(t_3)\right]\times\mathrm{E}\left[B(t_2)-B(t_1)\right] \\ &= \left(\mathrm{E}\left[B(t_4)\right]-\mathrm{E}\left[B(t_3)\right]\right)\times\left(\mathrm{E}\left[B(t_2)\right]-\mathrm{E}\left[B(t_1)\right]\right) \\ &= (0-0)\times(0-0)=0\end{aligned}$$

The first line of the preceding equation is true because the two intervals are non-overlapping and therefore the quantities ($B(t_4) - B(t_3)$) and ($B(t_2) - B(t_1)$) are independent, due to the assumption of independent increments. Recall that for independent random variables X and Y, E[XY] = E[X]E[Y]. Now consider the two intervals (0, t_1) and (t_1, t_2), for $0 < t_1 < t_2$. These are nonoverlapping intervals, so

$$\begin{aligned}0=&\mathrm{E}\left[\left(B(t_2)-B(t_1)\right)\left(B(t_1)-B(0)\right)\right] \\ &= \mathrm{E}\left[\left(B(t_2)-B(t_1)\right)B(t_1)\right] \\ &= \mathrm{E}\left[B(t_2)B(t_1)\right]-\mathrm{E}\left[B^2(t_1)\right] \\ &= \mathrm{E}\left[B(t_2)B(t_1)\right]-\mathrm{Var}\left[B(t_1)\right] \\ &= \mathrm{E}\left[B(t_2)B(t_1)\right]-t_1\end{aligned}$$

Therefore,

$R_B(t_1, t_2)$ = E[$B(t_1)B(t_2)$] = t_1 where $t_1 < t_2$

In general, then, the autocorrelation of $B(t)$ can be expressed as $R_B(t, s) = \min [t, s]$. Because $B(t)$ has zero mean, the autocovariance is the same as the autocorrelation. Thus $C_B(t, s) = \min [t, s]$.

For any $t \geq 0$ and $\delta > 0$, the increment of a Brownian motion process, $B(t + \delta) - B(t)$, is normally distributed with mean 0 and variance δ. Thus,

$$\Pr\left[\left(B(t+\delta)-B(t)\right) \leq x\right] = \frac{1}{\sqrt{2\pi\delta}} \int_{-\infty}^{x} e^{-y^2/2\delta}\, dy \qquad \textbf{(19.5)}$$

Note that this distribution is independent of t and depends only on δ, consistent with the fact that $B(t)$ has stationary increments.

One useful way to visualize the Brownian motion process is as the limit of a discrete-time process. Here, we follow the development in [FALC90]. Let us consider a particle performing a random walk on the real line. At small time intervals τ, the particle randomly jumps a small distance δ to the left or right. We denote the position of the particle at time $k\tau$ as $X_\tau(k\tau)$. If positive and negative jumps are equally likely, then $X_\tau((k + 1)\tau)$ equals $X_\tau(k\tau) + \delta$ or $X_\tau(k\tau) - \delta$ with equal probability. If we assume that $X_\tau(0) = 0$, then the position of the particle at time t is

$$X_\tau(t) = \delta\,(Y_1 + Y_2 \ldots + Y_{\lfloor t/\tau \rfloor})$$

where Y_1, Y_2 are independent random variables with equal probability of being 1 or –1 and $\lfloor t/\tau \rfloor$ denotes the largest integer less than or equal to t/τ. It is convenient to normalize the step length δ as $\sqrt{\tau}$ so that

$$X_\tau(t) = \sqrt{\tau}\,(Y_1 + Y_2 \ldots + Y_{\lfloor t/\tau \rfloor})$$

By the central limit theorem, for fixed t, if τ is sufficiently small then the sum in the preceding equation consists of many random variables and therefore the distribution of $X_\tau(t)$ is approximately normal with mean 0 and variance t, because the Y_i have mean 0 and variance 1. Also, for fixed t and h, if τ is sufficiently small, then $X_\tau(t + h) - X_\tau(t)$ is approximately normal with mean 0 and variance h. Finally, we note that the increments of $X_\tau(t)$ are independent. Thus, $X_\tau(t)$ is a discrete-time function that approximates Brownian motion. If we divide the time axis more finely, we improve the approximation. In the limit, this becomes a continuous-time Brownian motion process.

POISSON AND RELATED PROCESSES

Recall that for random arrivals in time, we have the Poisson distribution:

$$\Pr\left[k \text{ items arrive in time interval } T\right] = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$$

We can define a **Poisson counting process** $\{N(t), t \geq 0\}$ as follows:

1. $N(t)$ has stationary independent increments.
2. $N(0) = 0$.
3. For $0 < t_1 < t_2$, the quantity $N(t_2) - N(t_1)$ equals the number of points in the interval (t_1, t_2) and is Poisson distributed with mean $\lambda(t_1 - t_2)$.

Then we have the following probability functions for $N(t)$:

$$\Pr\left[N(t) = k\right] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

$$\mathrm{E}\left[N(t)\right] = \mathrm{Var}\left[N(t)\right] = \lambda t$$

Clearly, $N(t)$ is not stationary, because its mean is a function of time. Every time function of this stochastic process (one outcome) has the form of an increasing staircase with steps equal to 1, occurring at the random points t_i. Figure 19.4a gives an example of $N(t)$ for a specific outcome.

A stationary process related to the Poisson counting process is the **Poisson increment process**. For a Poisson counting process $N(t)$ with mean λt, and for a constant L ($L > 0$), we can define the Poisson increment process $X(t)$ as follows:

$$X(t) = \frac{N(t+L) - N(t)}{L}$$

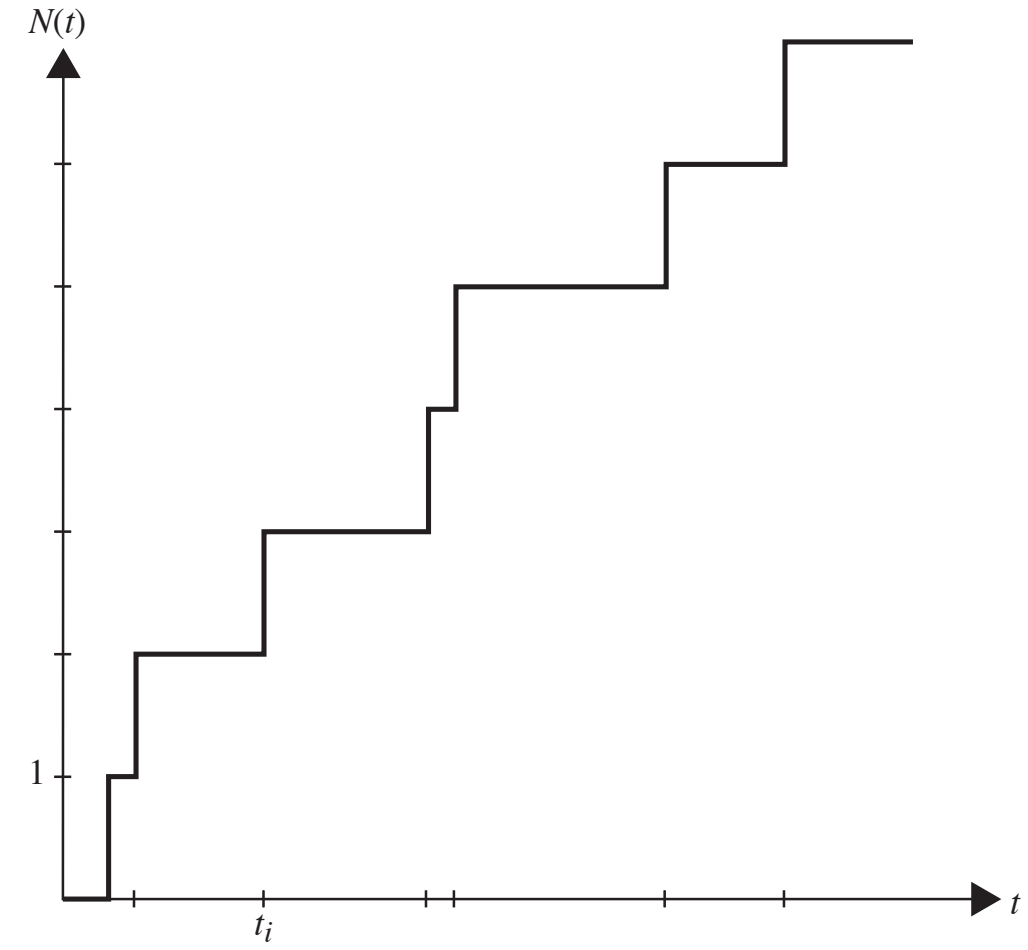
$X(t)$ equals k/L, where k is the number of points in the interval $(t, t + L)$. The increment process derived from the counting process in Figure 19.4a is shown in Figure 19.4b. The following relationship holds.

$$\mathrm{E}[X(t)] = \frac{1}{L}\mathrm{E}[N(t+L)] - \frac{1}{L}\mathrm{E}[N(t)] = \lambda$$

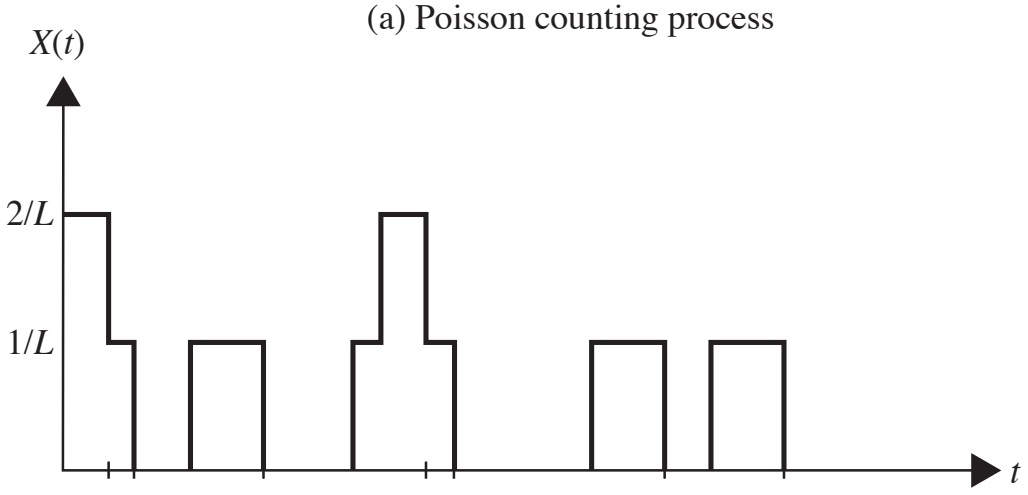
With a constant mean, $X(t)$ is a wide-sense stationary process and therefore has an autocorrelation function of a single variable, $R(\tau)$. It can be shown that this function is:

$$R(\tau) = \begin{cases} \lambda^2 & |\tau| > L \\ \lambda^2 + \dfrac{\lambda^2}{L}\left(1 - \dfrac{|\tau|}{L}\right) & |\tau| < L \end{cases} \qquad \textbf{(19.6)}$$

Thus, the correlation is greatest if the two time instants are within the interval length of each other, and it is a small constant value for greater time differences.



(a) Poisson counting process



(b) Poisson increment process

Figure 19.4 Poisson Processes

Ergodicity

For a stochastic process $\mathbf{x}(t)$, there are two types of "averaging" functions that can be performed: ensemble averages and time averages.

First, consider **ensemble averages**. For a constant value of t, $\mathbf{x}(t)$ is a single random variable with a mean, variance, and other distributional properties. For a given constant value C of t, the following measures exist:

$$\mathrm{E}\left[\mathbf{x}(\mathrm{C})\right] = \mu_{\mathbf{x}}(\mathrm{C}) = \int_{-\infty}^{\infty} x f(x;\mathrm{C})\,dx \qquad \text{continuous-value case}$$

$$\mathrm{E}\left[\mathbf{x}(\mathrm{C})\right] = \mu_{\mathbf{x}}(\mathrm{C}) = \sum_{\text{all } k} k \Pr\left[\mathbf{x}(\mathrm{C}) = k\right] \qquad \text{discrete-value case}$$

$$\mathrm{Var}\left[\mathbf{x}(\mathrm{C})\right] = \sigma^2_{\mathbf{x}(\mathrm{C})} = \mathrm{E}\left[\left(\mathbf{x}(\mathrm{C}) - \mu_{\mathbf{x}}(\mathrm{C})\right)^2\right] = \mathrm{E}\left[\mathbf{x}(\mathrm{C})^2\right] - \mu^2_{\mathbf{x}}(\mathrm{C})$$

Each of these quantities is calculated over all values of $\mathbf{x}(t)$ for all possible outcomes. For a given random variable, the set of all possible outcomes is called an *ensemble*, and hence these are referred to as ensemble averages.

For time averages, consider a single outcome of $\mathbf{x}(t)$. This is a single deterministic function of t. Looking at $\mathbf{x}(t)$ in this way, we can consider what is the average value of the function over time. This **time average** is generally expressed as follows:

$$M_T = \frac{1}{2T}\int_{-T}^{T} \mathbf{x}(t)\,dt \qquad \text{continuous-time case}$$

$$M_T = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}(t) \qquad \text{discrete-time case}$$

Note that M_T is a random variable, because the calculation of M_T for a single time function is a calculation for a single outcome.

A stationary process is said to be **ergodic** if time averages equal ensemble averages. Because E[**x**(t)] is a constant for a stationary process, we have

$$\mathrm{E}[M_T] = \mathrm{E}[\mathbf{x}(t)] = \mu$$

Thus we can say that a stationary process is ergodic if

$$\lim_{T \to \infty} \mathrm{Var}(M_T) = 0$$

In words, as the time average is taken over larger and larger time intervals, the value of the time average approaches the ensemble average.

The conditions under which a stochastic process is ergodic are beyond the scope of this book, but the assumption is generally made. Indeed, the assumption of ergodicity is essential to almost any mathematical model used for stationary stochastic processes. The practical importance of ergodicity is that in most cases, one does not have access to the ensemble of outcomes of a stochastic process or even to more than one outcome. Thus, the only means of obtaining estimates of the probabilistic parameters of the stochastic process is to analyze a single time function over a long period of time.

19.4 RECOMMENDED READING

There is a huge collection of books, dating back to the seventeenth century, on probability and random processes. My personal favorite is [HAMM91]; it is both a very practical book on the application of probability and an enlightening treatment of the philosophy of probability. A good book for self-

study is [GOLD87]; it contains numerous problems with worked-out solutions. Another good, comprehensive treatment is [ROSS10].

There are also many books on stochastic processes. My personal favorite is [PAPO02]; this book has remained consistently excellent through multiple editions stretching back to 1965. Another good book, suitable for self-study, is [GRIM01a, GRIM01b]; the second volume contains numerous exercises with solutions.

GOLD87 Goldberg, S. *Probability: An Introduction.* New York: Dover, 1987.

GRIM01a Grimmett, G., and Stirzaker, D. *Probability and Random Processes.* Oxford: Oxford University Press, 2001.

GRIM01b Grimmett, G., and Stirzaker, D. *One Thousand Exercises in Probability.* Oxford: Oxford University Press, 2001.

HAMM91 Hamming, R. *The Art of Probability: For Scientists and Engineers.* Reading, MA: Addison-Wesley, 1991.

PAPO02 Papoulis, A., and Unnikrishna, P. *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, 2002.

ROSS10 Ross, S. *First Course in Probability.* Upper Saddle River, NJ: Prentice Hall, 2010.

19.5 PROBLEMS

19.1 You are asked to play a game in which I hide a prize in one of three boxes (with equal probability for all three boxes) while you are out of the room. When you return, you have to guess which box hides the prize. There are two stages to the game. First, you indicate one of the three boxes as your choice. As soon as you do that, I open the lid of one of the other two boxes and I will always open an empty box. I can do this because I know where the prize is hidden. At this point, the prize must be in the box that you have chosen or in the other unopened box. You are now free to stick with your original choice or to switch to the other unopened box. You win the prize if your final selection is the box containing the prize. What is your best strategy?

Should you (a) stay with your original choice, (b) switch to the other box, or (c) do either because it does not matter?

19.2 A patient has a test for some disease that comes back positive (indicating he has the disease). You are told that

- the accuracy of the test is 87% (i.e., if a patient has the disease, 87% of the time, the test yields the correct result, and if the patient does not have the disease, 87% of the time, the test yields the correct result)
- the incidence of the disease in the population is 1%

Given that the test is positive, how probable is it that the patient really has the disease?

19.3 A taxicab was involved in a fatal hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are told that:

- 85% of the cabs in the city are Green and 15% are Blue
- A witness identified the cab as Blue

The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness was correct in identifying the color of the cab 80% of the time. What is the probability that the cab involved in the incident was Blue rather than Green?

19.4 The birthday paradox is a famous problem in probability that can be stated as follows: What is the minimum value of K such that the probability is greater than 0.5 that at least two people in a group of K people have the same birthday? Ignore February 29 and assume that each birthday is equally likely. We will do the problem in two parts.

a. Define $Q(K)$ as the probability that there are no duplicate birthdays in a group of K people. Derive a formula for $Q(K)$. *Hint:* First determine the number of different ways, N, that we can have K values with no duplicates.

b. Define $P(K)$ as the probability that there is at least one duplicate birthday in a group of K people. Derive this formula. What is the minimum value of K such that $P(K) > 0.5$? It may help to plot $P(K)$.

19.5 A pair of fair dice (the probability of each outcome is 1/6) is thrown. Let X be the maximum of the two numbers that comes up.

a. Find the distribution of X.

b. Find the expectation E[X], the variance Var[X], and the standard deviation σ_X.

19.6 A player tosses a fair die. If a prime number greater than 1 appears, he wins that number of dollars, but if a nonprime number appears, he loses that number of dollars.

a. Denote the player's gain or loss on one toss by the random variable X. Enumerate the distribution of X.

b. Is the game fair (i.e., E[X] = 0)?

19.7 In the carnival game known as *chuck-a-luck*, a player pays an amount E as an entrance fee, selects a number between one and six, and then rolls three dice. If all three dice show the number selected, the player is paid four times the entrance fee; if two dice show the number, the player is paid three times the entrance fee; and if only one die shows the number, the player is paid twice the entrance fee. If the selected number does not show up, the player is paid nothing. Let X denote the player's gain in a single play of this game, and assume the dice are fair.

a. Determine the probability function of X.

b. Compute E[X].

19.8 The mean and variance of X are 50 and 4, respectively. Evaluate

a. the mean of X^2

b. the variance and standard deviation of $2X + 3$

c. the variance and standard deviation of $-X$

19.9 The continuous random variable R has a uniform density between 900 and 1,100, and zero elsewhere. Find the probability that R is between 950 and 1,050.

19.10 Show that, all other things being equal, the greater the correlation coefficient of two random variables, the greater the variance of their sum and the less the variance of their difference.

19.11 Suppose that X and Y each have only two possible values, 0 and 1. Prove that if X and Y are uncorrelated, then they are also independent.

19.12 Consider a random variable X with the following distribution: Pr[$X = -1$] = 0.25; Pr[$X = 0$] = 0.5; Pr[$X = 1$] = 0.25. Let $Y = X^2$.

a. Are X and Y independent random variables? Justify your answer.

b. Calculate the covariance Cov(X, Y).

c. Are X and Y uncorrelated? Justify your answer.

19.13 An artificial example of a stochastic process is a deterministic signal $\mathbf{x}(t) = g(t)$. Determine the mean, variance, and autocorrelation of $\mathbf{x}(t)$.

19.14 Suppose that $\mathbf{x}(t)$ is a stochastic process with

$$\mu(t) = 3 \quad R(t_1, t_2) = 9 + 4e^{-0.2|t_1 - t_2|}$$

Determine the mean, variance, and covariance of the following random variables $Z = \mathbf{x}(5)$ and $W = \mathbf{x}(8)$.

19.15 Let $\{\mathbf{Z_n}\}$ be a set of uncorrelated real-valued random variables, each with a mean of 0 and a variance of 1. Define the moving average

$$\mathbf{Y_n} = \sum_{i=0}^{K} \alpha_i \mathbf{Z_{n-i}}$$

for constants $\alpha_0, \alpha_1, \ldots, \alpha_K$. Show that $\mathbf{Y}$ is stationary and find its autocovariance function.

19.16 Let $\mathbf{X_n} = \mathbf{A}\cos(n\lambda) + \mathbf{B}\sin(n\lambda)$ where $\mathbf{A}$ and $\mathbf{B}$ are uncorrelated random variables, each with a mean of 0 and a variance of 1. Show that $\mathbf{X}$ is stationary with a spectrum containing exactly one point.