2

Brief Review of Probabilistic Modeling

Our focus in this chapter is on the construction of models, given descriptions of physical situations. Since many of these models will contain elements of uncertainty, at least from the model builder's point of view, we should review briefly the fundamentals of probabilistic modeling. While this material will serve as a review for those readers who have a solid grounding in probabilistic reasoning, it is still recommended reading for all, since concepts and results are developed that are of direct use in later chapters.

2.1 EXPERIMENT, SAMPLE SPACE, AND EVENTS

Each probabilistic situation that we wish to analyze can be viewed in the context of an experiment. By *experiment* we mean any nondeterministic process that has a number of distinct possible outcomes. Thus, an experiment is first characterized by a list of possible outcomes. A particular performance of the experiment, sometimes referred to as an experimental *trial*, yields one and only one of the outcomes.

The finest-grained list of outcomes for an experiment is the *sample space* of the experiment. Examples of sample spaces are:

1. {heads, tails} in the simple toss of a coin.
2. {1, 2, 3, . . .}, describing the possible number of fire alarms in a city during a year.
3. $\{0 \leq x \leq 10, 0 \leq y \leq 10\}$, describing the possible locations of required on-the-scene social services in a city 10 by 10 miles square.

As these three examples indicate, the number of *elements* or *points* in a sample space can be finite, countably infinite, or noncountably infinite. Also, the elements may be something other than numbers (e.g., "heads" or "tails").

Probabilistic analysis requires considerable manipulation in an experiment's sample space. For this, we require knowledge of the algebra of events, where an *event* is defined to be a collection of points in the sample space. A generic event is given an arbitrary label, such as A, B, or C. Since the entire sample space defines the universe of our concerns, it is called U, for *universal event*. An event containing no points in the sample space is called $\varnothing$, the *empty event* (or null set). There are three key operations in the algebra of events:

1. **Union.** $A \cup B =$ set of all points in *either* A or B.
2. **Intersection.** $A \cap B =$ set of all points in *both* A and B.
3. **Complement.** $A' =$ set of all points (in U) *not* in A.

These three operations are governed by the following seven algebraic axioms:

1.	$A \cup B = B \cup A$	Commutative law.
2.	$A \cup (B \cup C) = (A \cup B) \cup C$	Associative law.
3.	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	Distributive law.
4.	$(A')' = A$	Complement of the complement of an event is the original event.
5.	$(A \cap B)' = A' \cup B'$	Complement of the intersection of two events is the union of their complements.
6.	$A \cap A' = \varnothing$	Intersection of an event with its complement is the empty event.
7.	$A \cap U = A$	Intersection of an event with the universe of events is the original event.

The events $A_1, A_2, \ldots, A_N$ are said to be *mutually exclusive* if they share no point(s) in common:

$$A_i \cap A_j = \begin{cases} A_i & \text{if } i = j \\ \varnothing & \text{if } i \neq j \end{cases} \qquad i, j = 1, 2, \ldots, N$$

Events $A_1, A_2, \ldots, A_N$ are said to be *collectively exhaustive* if they include all points in the universe of events:

$$\bigcup_{i=1}^{N} A_i = U$$

A set of events $A_1, A_2, \ldots, A_N$ that are both mutually exclusive and collectively exhaustive contains each point of the sample space U in one and only one of the events A_i.

Given these notions, we can more carefully define a *sample space* as follows:

> *Definition:* A *sample space* is the finest-grained, mutually exclusive, collectively exhaustive listing of all possible outcomes of an experiment.

The first (and perhaps most important) step in constructing a probabilistic model consists of identifying the sample space for the corresponding experiment.

Example 1: Stick Cutting

A simple example drawn from geometrical probability will serve to illustrate some of these ideas. Suppose that two points are marked (in some nondeterministic way) on a stick of length 1 meter.

a. Define the sample space for this experiment.

b. Identify the event, "The second point is to the left of the first point."

c. Suppose that the stick is cut at the marked points. Identify the event, "A triangle can be formed with the resulting three pieces."

Solution:

a. Call the first point x_1 and the second x_2. Since we are given no information about x_1 and x_2 other than that each is between 0 and 1, the sample space is the collection of points in the unit square shown in Figure 2.1.

b. The event indicated, call it E_1, corresponds to $(x_1 > x_2)$. This set of points lies in the triangular region of the sample space below the line $x_2 = x_1$.

c. Let Δ be the event that a triangle can be formed. Identification of Δ in the (x_1, x_2) sample space requires a bit more care than in the case of the event $E_1 = (x_1 > x_2)$. Suppose that, in fact, $x_1 > x_2$. Then the three stick lengths are x_2, $x_1 - x_2$, and $1 - x_1$. For a triangle to be formed, each length must be less than $\frac{1}{2}$:

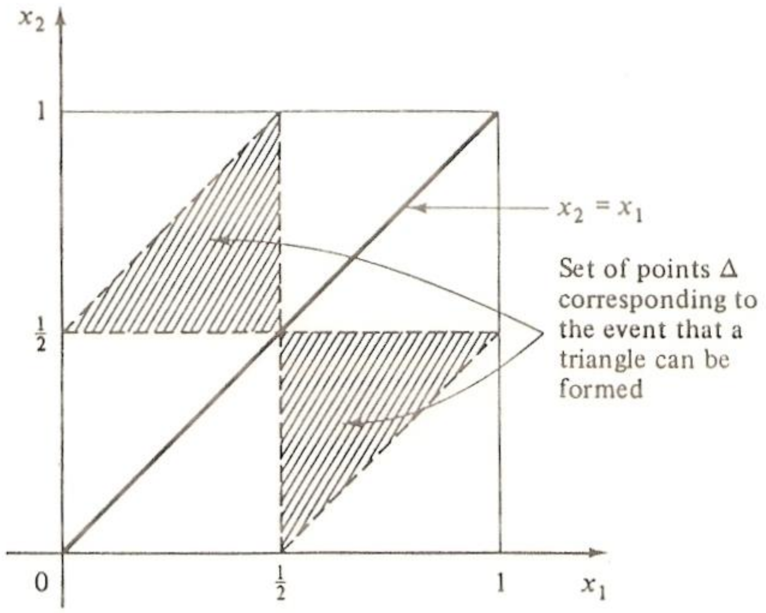


FIGURE 2.1 Sample space for the broken-stick experiment.

$$A: \quad x_2 < \tfrac{1}{2}$$
$$B: \quad x_1 - x_2 < \tfrac{1}{2}$$
$$C: \quad 1 - x_1 < \tfrac{1}{2}$$

So, given $x_1 > x_2$, Δ is composed of the set of points that simultaneously satisfies these three inequalities or, equivalently, the set of points in $A \cap B \cap C$. The resulting set is contained in the lower triangle of area $\frac{1}{8}$ in the sample space. Now suppose that $(x_1 \leq x_2) = E_1'$. Here we can invoke *symmetry*. A current dictionary defines symmetry as "similarity of form or arrangement on either side of a dividing line or plane; correspondence of opposite parts in size, shape, and position."[1] Since the labeling of points x_1 and x_2 was totally arbitrary, there exists symmetry about the line $x_1 = x_2$. Thus, we obtain a similar triangle of area $\frac{1}{8}$ above the line $x_1 = x_2$.

As one final point, we may wish to express Δ in terms of the algebra of events. If we define

$$D: \quad x_1 < \tfrac{1}{2}$$
$$E: \quad x_2 - x_1 < \tfrac{1}{2}$$
$$F: \quad 1 - x_2 < \tfrac{1}{2}$$

then

$$\Delta = [(A \cap B \cap C) \cap E_1] \cup [(D \cap E \cap F) \cap E_1']$$

1 *Webster's New World Dictionary of the English Language*, Second College Edition, World Publishing Co., New York, 1974, p. 1442.

In modeling experiments, extreme care must be given to a precise interpretation of the word statement. Statements that may at first sound the same may actually imply markedly different experiments; or, statements may simply be imprecise and ambiguous. A famous illustration of this in a geometrical setting, known as *Bertrand's paradox* (1907), yields three different answers to the question: What is the probability that a "random chord" of a circle of unit radius has a length greater than $\sqrt{3}$, the side of an inscribed equilateral triangle?[2] Each of the three solutions, which we will develop in Chapter 3, is "correct" since each involves a different interpretation of that difficult word "random."

To strengthen your understanding of word statements as they relate to concepts of sample space and event, try the following two exercises:

> **Exercise 2.1: Stick Breaking, mod 2** Repeat parts (a) and (c) of the triangle problem described in Example 1, given that the problem statement is changed as follows: "A point is marked (in some random way) on a stick of length 1 meter. Then a second point is marked on the stick (in some random way) to the left of the first point."

> **Exercise 2.2: Stick Breaking, Still Again!** Repeat parts (a), (b), and (c) of the triangle problem given that the problem statement is changed as follows: "A point is marked (in some random way) on a stick of length 1 meter. The stick is then cut at that point. Another point is marked (in some random way) on the larger of the two resulting stick pieces."

2.2 EVENT PROBABILITIES

The second step in constructing a probabilistic model is to assign probabilities to events in the sample space. For any arbitrary event A, we say that $P\{A\}$ is the probability that an outcome of the experiment is contained or included in event A. This is a clearer statement than saying that $P\{A\}$ is the "probability of event A occurring," a statement that sometimes leads to confusion. Unless event A is a single point in the sample space, event A never "occurs" in its totality, but rather a single element or point in A may be the outcome of a particular experiment. The assigned event probabilities must obey the three axioms of probability:

1. For any event A, $P\{A\} \geq 0$ (nonnegativity of probabilities).
2. $P\{U\} = 1$ (totality of the universe U).
3. If $A \cap B = \varnothing$, then $P\{A \cup B\} = P\{A\} + P\{B\}$ (additivity of probabilities of mutually exclusive events).

[2] J. Bertrand, *Calcul des probabilités*, Paris, 1907.

For the moment we will not be concerned with the method of assigning probabilities. We will assume that these probabilities are assigned to each finest-grained outcome of the experiment so that for each event A one can compute $P\{A\}$ by simply summing the probabilities of the finest-grained outcomes comprising A. As we will see shortly, this summation could entail the sum of a finite number of elements or a countably infinite number of elements (in a countably infinite sample space), or it could entail integration (in a noncountably infinite sample space).

Sometimes we have conditioning events in a sample space, reflecting partial information about the experimental outcome, and we wish to know what this means about the likelihood of other events "occurring or not occurring," given the conditioning event. Thus, we define *conditional probability* as

$$P\{B \mid A\} = P\{\text{outcome of the experiment is contained in event } B, \text{ given that it is contained in event } A\}$$

Given that the conditioning event requires that the collection of "B-type" outcomes that could occur must also be contained in A, we could rewrite the definition of conditional probability as

$$P\{B \mid A\} = P\{\text{outcome of the experiment is contained in event } B \cap A, \text{ given that it is contained in event } A\}$$

In manipulating conditional probabilities, the set of outcomes contained in the conditioning event A now constitutes the universal set of outcomes. Where "before the fact" (of A) the a priori universe was U, "after the fact" the a posteriori universe is A. Given the conditioning event, the new universe A is to be treated just as a sample space. Thus, the probabilities distributed over the finest-grained outcomes in A must be scaled so that their total (conditional) probability sums to 1. To do this, any event C that is fully contained in A (i.e., $A \cap C = C$ or $A \cup C = A$) must have its corresponding probability scaled by $1/P\{A\}$. Thus,

$$P\{C \mid A\} = \frac{P\{C\}}{P\{A\}} \qquad \text{(assuming that } C \text{ is contained in } A\text{)}$$

Since $A \cap B$ is the collection of all outcomes in both A and B, it must be true that $A \cap B$ is contained in A, and thus

$$P\{B \mid A\} = \frac{P\{A \cap B\}}{P\{A\}} \qquad \text{where } P\{A\} > 0 \tag{2.1}$$

This is the operational definition of conditional probability.

When dealing with the intersection of events, we will on occasion substitute a comma for the intersection operator. As an example, given two conditioning events A_1 and A_2, $P\{B | A_1 \cap A_2\}$ and $P\{B | A_1, A_2\}$ have the same meaning: the probability that the experimental outcome is contained in event B, given that it is contained in *both* A_1 and A_2.

Two events are said to be *independent* if information concerning the occurrence of one of them does not alter the probability of occurrence of the other. Formally, events A and B, with $P\{A\} > 0$ and $P\{B\} > 0$, are said to be independent if and only if

$$P\{B | A\} = P\{B\}$$

Using the definition of conditional probability, we can write

$$P\{A \cap B\} = P\{A\}P\{B | A\} = P\{B\}P\{A | B\}$$

Thus, independence implies that $P\{A | B\} = P\{A\}$ and that

$$P\{A \cap B\} = P\{A\}P\{B\} \tag{2.2}$$

Question: If A and B are mutually exclusive, can they be independent?

Question: If A and B are collectively exhaustive, can they be independent?

Exercise 2.3: Independence of Events If A and B are independent, show that A and B' are independent, as are A' and B, and A' and B'.

Suppose that we have a collection of N events, $A_1, A_2, \ldots, A_N$. These events are said to be *mutually independent* if

$$P\{A_i | A_{n_1}, \ldots, A_{n_p}\} = P\{A_i\} \qquad \text{for all } i \neq n_1, n_2, \ldots, n_p;$$
$$n_j = 1, 2, \ldots, N; \quad p = 1, 2, \ldots, N-1.$$

In other words, for each possible event A_i, information on the occurrence of any combination of the other events does not affect the probability that the experimental outcome is contained in event A_i. It is important to be aware that events may be pairwise independent or be otherwise conditionally independent but not mutually independent. Only with mutual independence does "information about the 'other' events A_j, $j \neq i$, tell us nothing about event A_i."

2.3 RANDOM VARIABLES

Many, if not most, experiments have numerical values associated with different outcomes. In fact, for some experiments the finest-grained outcomes are described directly in terms of numbers. This is the case for the triangle problem, but not for a simple flip of a coin (in which the two possible outcomes are "heads" or "tails"). But even for the flip of the coin we may wish to associate numerical values with the outcomes, say +\$1 for "heads" and −\$2 for "tails." To facilitate such numerical descriptions of the outcomes of an experiment, we introduce the notion of a random variable.

Definition: Given an experiment with a sample space and a probability assignment over the sample space, a *random variable* is a function that assigns a numerical value to each finest-grained outcome in the sample space.

Note: While each finest-grained outcome is unique, it is not necessary that each assigned value of the random variable be unique. Thus, two or more finest-grained points may yield the same value of the random variable.

We shall usually denote random variables by capital letters, such as X, Y, or Z. Each of these represents a complete set of correspondences between finest-grained outcomes in the experiment, with their probability assignments, and associated numerical values. Thus, the notation X does not refer to a number such as 2.3 or $-\pi$, but rather to a listing (or, if you like, mapping) which provides a numerical value for the random variable for each point in the sample space. This sample space is assumed to have a probability assignment associated with it. Once an experiment is carried out, each particular outcome yields specific numerical values for the random variables, say $X = x$, $Y = y$, and $Z = z$. In general and whenever convenient, we will use the same letter (but in lowercase form) to indicate a particular experimental value of the random variable. However, it would also be entirely reasonable to say that $X = 2.62$, $X = a$, $X = c + \pi/3$, or $X =$ any other number or representation for a number.

The set of possible values for a random variable is called its *event space*. For the purpose of summing probabilities, it is convenient to discuss separately *discrete random variables*, whose event spaces contain a finite or countably infinite number of values, and *continuous random variables*, whose event spaces contain a noncountably infinite number of values.

2.4 PROBABILITY MASS FUNCTION

We define for a discrete random variable X,

$p_X(x) \equiv$ probability that the random variable X assumes the experimental value x on a performance of the experiment

or, in shorthand,

$$p_X(x) \equiv P\{X = x\}$$

The function $p_X(\cdot)$ is the *probability mass function* (pmf) of the (discrete) random variable X. Clearly, we must have

$$\sum_x p_X(x) = 1$$

$$0 \leq p_X(x) \leq 1 \qquad \text{for all } x$$

The probability mass function is the assignment of probabilities to each possible value of the random variable. It plays an identically analogous role for random variables that the assignment of probabilities to finest-grained outcomes plays in the original sample space.

Associated with each pmf is its *cumulative distribution function* (cdf), which is simply defined to be the probability that the random variable assumes an experimental value less than or equal to a specified amount,

$$P_X(x) \equiv P\{X \leq x\} = \sum_{\text{all } y \leq x} p_X(y) \tag{2.3}$$

Note that the cdf for a discrete random variable is a step function. It starts at zero for x values less than the smallest possible and proceeds from left to right in steps, the height of a step at $X = x_0$ equaling the probability that X will assume that particular experimental value. The step function eventually reaches unity as a maximum value since $P_X(\infty) = 1$.

If we wish to obtain information about two or more (discrete) random variables simultaneously, we must introduce the concept of *compound* (or *joint*) probability mass functions. For instance, for two random variables X and Y, their compound pmf is given by

$$p_{X,Y}(x, y) = P\{X = x, \ Y = y\} \qquad \text{all } x, y$$

In general, given N discrete random variables $X_1, X_2, \ldots, X_N$, there exists a corresponding N-argument pmf,

$$p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) = P\{X_1 = x_1, X_2 = x_2 \ldots, X_N = x_N\}$$

Clearly,

$$0 \leq p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) \leq 1$$

$$\sum_{\text{all } x_1,x_2,\ldots,x_N} p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) = 1$$

In cases involving multiple random variables $X_1, X_2, \ldots, X_N$,

$$p_{X_i}(x) \equiv P\{X_i = x\}$$

is said to be the *marginal pmf* for X_i. We can calculate the marginal from the point pmf simply by summing over all the values of the other random variables:

$$p_{X_i}(x) = \sum_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N} p_{X_1, X_2, \ldots, X_N}(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_N) \quad (2.4)$$

2.5 CONDITIONAL PMF'S AND INDEPENDENCE

Suppose we are told that an experimental outcome is contained in event A. We then wish to explore the probabilistic behavior of random variables X and Y, given A. Following the definition of conditional probability, we introduce the *conditional compound pmf*,

$$p_{X,Y}(x, y \mid A) = \begin{cases} \dfrac{p_{X,Y}(x, y)}{P\{A\}} & (x, y) \in A, \quad P\{A\} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.5)$$

If event A is stated in terms of the specific experimental value of one of the random variables of the experiment, we introduce another notation:

$$p_{X|Y}(x \mid y) \equiv P\{X = x \mid Y = y\}$$

By conditional probability,

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \qquad p_Y(y) > 0 \quad (2.6)$$

These notations and definitions extend in an obvious way to situations with more than two random variables.

Example 2: Minibus

Suppose that a minibus with capacity for five passengers departs from a commuter station. Observation has shown that the bus never departs empty (with no passengers) but that each possible positive number of passengers is equally likely to be on the bus at departure time. Passengers are of two types: male and female. Given that the departing bus contains exactly n passengers ($n = 1, 2, \ldots, 5$), each possible combination of male and female passengers has been found to be equally likely.

a. Identify the sample space and joint probability mass function for this experiment.

b. Determine the marginal pmf for the number of females on the minibus.

c. Determine the joint conditional pmf for the number of females and the number of males on the minibus, given that the bus departs at full capacity.

d. Determine the conditional marginal pmf for the number of females, given that there are at least twice as many females as males on the minibus.

Solution:

There are two random variables in this experiment:

$$N_F \equiv \text{number of females on the minibus}$$
$$N_M \equiv \text{number of males on the minibus}$$

a. A complete listing of their possible (paired) values constitutes the sample space for this experiment. These points occupy the nearly triangular region shown in Figure 2.2. The region is not perfectly triangular since the origin ($N_F = 0$, $N_M = 0$) is excluded because the minibus never departs empty. The region is bounded above by the line $n_F + n_M = 5$, which expresses the capacity constraint for the minibus.

To determine the joint pmf for N_F and N_M, we must use the conditional information given in the word statement. We know that any particular positive total number of passengers, ranging up to 5, is equally likely. Let

$$A_i = \text{event that } i \text{ total passengers are on the bus}$$

We know that $P\{A_i\} = \frac{1}{5}$, $i = 1, 2, \ldots, 5$. Points in an event A_i lie on the line $n_F + n_M = i$, as shown for A_3 in Figure 2.2. Given that an outcome of the experiment is contained in event A_i, we know that each of the points in A_i is equally likely. Since the number of points in A_i is equal to $i + 1$, we have

$$P\{N_F = n_F, N_M = n_M \mid A_i\} = \frac{1}{i+1} \qquad \text{for all positive integer } n_F, n_M \text{ such that } n_F + n_M = i$$

For any $\{n_F, n_M\}$ such that $n_F + n_M = i$, we can write

$$\begin{aligned} p_{N_F,N_M}(n_F, n_M) &= P\{N_F = n_F, N_M = n_M\} \\ &= \sum_{j=1}^{5} P\{N_F = n_F, N_M = n_M \mid A_j\}P\{A_j\} \\ &= P\{N_F = n_F, N_M = n_M \mid A_i\}P\{A_i\} \\ &= \frac{1}{i+1} \cdot \frac{1}{5} \qquad \text{for all positive integer } n_F, n_M \text{ such that } n_F + n_M = i. \end{aligned}$$

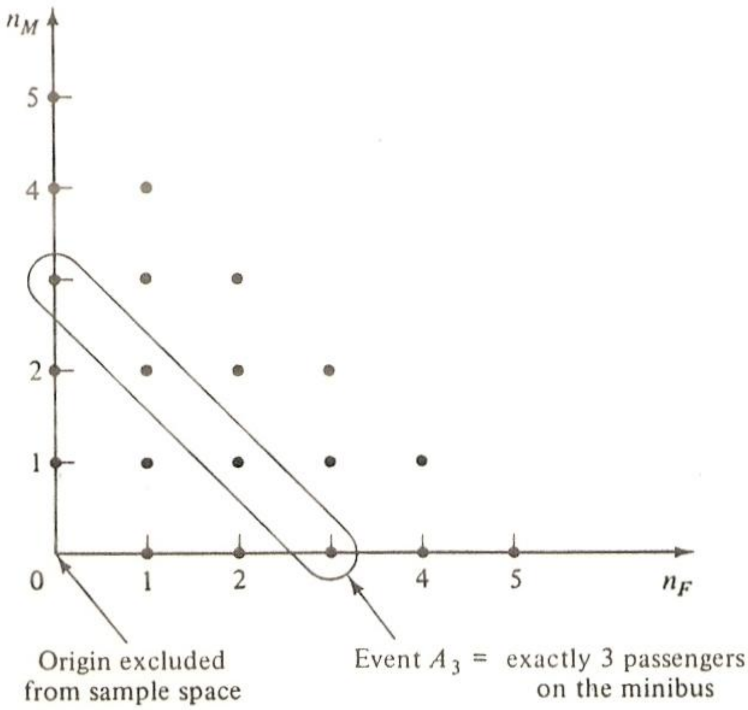


FIGURE 2.2 Sample space for minibus problem.

This is the answer to part (a). It says, roughly, that the probability that $\{N_F = n_F \textit{ and } N_M = n_M\}$ is equal to $\frac{1}{5}$ divided by 1 plus the sum $n_F + n_M$. For $i = 3$, for instance,

$$P\{N_F = 0, N_M = 3\} = P\{N_F = 1, N_M = 2\} = P\{N_F = 2, N_M = 1\}$$
$$= P\{N_F = 3, N_M = 0\} = \left(\frac{1}{3+1}\right) \cdot \frac{1}{5} = \frac{1}{20}$$

The complete joint pmf is shown in Figure 2.3.

b. Once we have the joint pmf for N_F and N_M, we can readily answer any question about the experiment. The marginal pmf for N_F is found by invoking (2.4), which simply asks us to sum over all values of N_M at each particular fixed value for N_F. For instance, to obtain $P\{N_F = 3\} = p_{N_F}(3)$, we sum the probabilities corresponding to the (finest-grained) events $\{N_F = 3, N_M = 0\}$, $\{N_F = 3, N_M = 1\}$, and $\{N_F = 3, N_M = 2\}$, yielding $\frac{1}{20} + \frac{1}{25} + \frac{1}{30} = \frac{37}{300}$. The complete pmf is shown in Figure 2.4.

c. If we are given conditional information that the bus departs at full capacity, we know that the experimental outcome is contained in event A_5 (i.e., $n_F + n_M = 5$). Thus, invoking (2.5),

$$p_{N_F,N_M}(n_F, n_M \mid A_5) = \begin{cases} \dfrac{\frac{1}{30}}{\frac{1}{5}} & (n_F, n_M) \in A_5 \\ 0 & \text{otherwise} \end{cases}$$

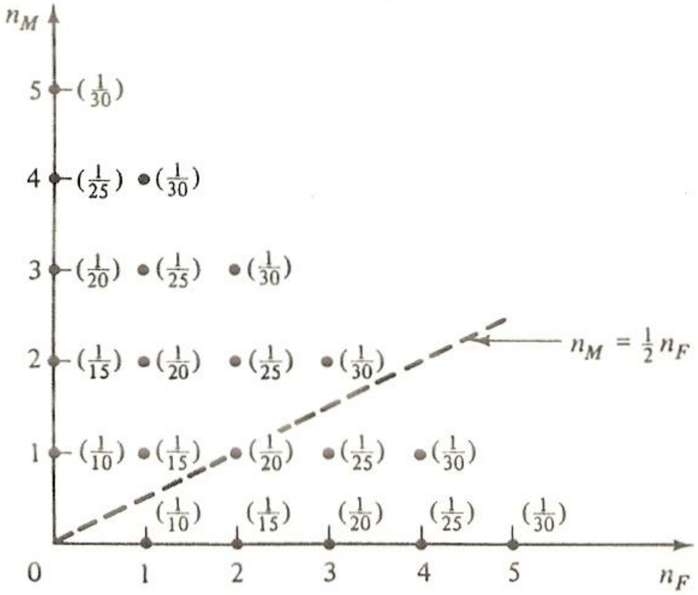


FIGURE 2.3 Joint pmf for minibus problem.

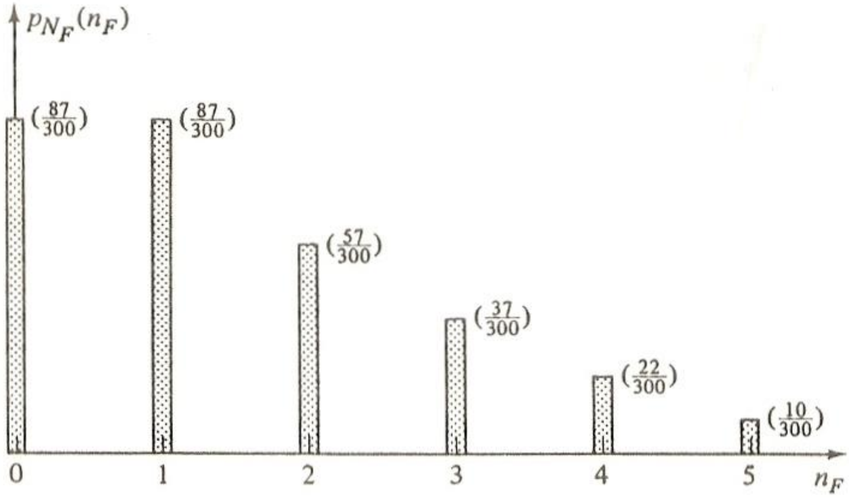


FIGURE 2.4 PMF for the number of female minibus passengers.

or

$$p_{N_F,N_M}(n_F, n_M \mid A_5) = \begin{cases} \frac{1}{6} & (n_F, n_M) \in A_5 \\ 0 & \text{otherwise} \end{cases}$$

This corresponds to a straight line of probability masses, each having mass $\frac{1}{6}$, at the integer points on the line $n_F + n_M = 5$ $(n_F, n_M \geq 0)$.

d. Let

B = event that "there are at least twice as many females as males on the minibus"

We want $p_{N_F}(n_F \mid B)$. First we work in the original joint sample space to determine finest-grained outcomes contained in the event B.

Clearly, these are points n_F, n_M satisfying the inequality $2n_M \leq n_F$. This corresponds to points lying on or below the line $n_M = \frac{1}{2} n_F$ (shown in Figure 2.3). Summing the probabilities of the eight finest-grained outcomes satisfying this inequality, we find that $P\{B\} = \frac{124}{300} = \frac{31}{75}$. Then, to find the conditional marginal pmf for N_F, given B, we simply sum the probabilities at a fixed value for n_F over all values of n_M contained in B, then scale by $1/\frac{31}{75}$. For instance,

$$p_{N_F}(2 \mid B) = \frac{\frac{1}{20} + \frac{1}{15}}{\frac{31}{75}} = \frac{15 + 20}{124} = \frac{35}{124}$$

The entire conditional marginal pmf is displayed in Figure 2.5. Notice how the conditional information has shifted the pmf for N_F toward greater numbers of females (compare to Figure 2.4).

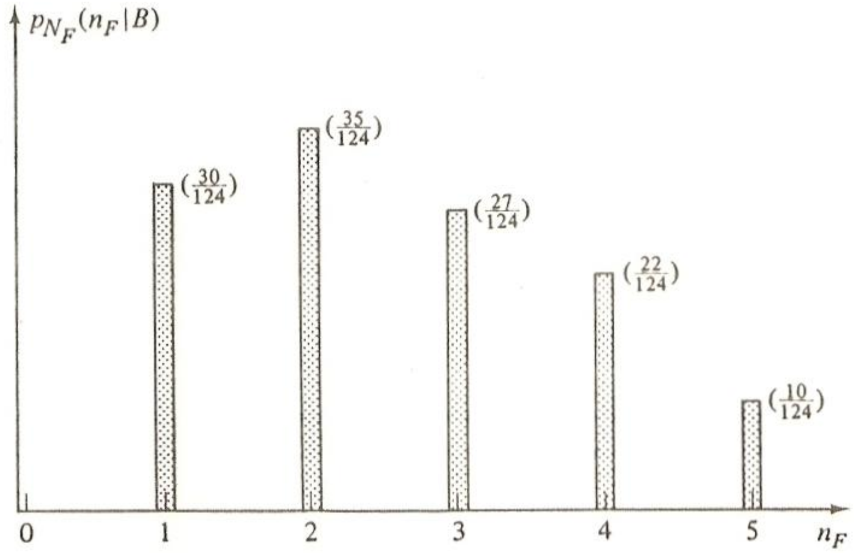


FIGURE 2.5 Conditional marginal pmf for the number of female passengers, given B.

Just as events can be independent, so, too, can random variables be independent. Intuitively, if X and Y are independent, any information regarding the value of one tells us nothing new about the value of the other. Formally, random variables X and Y are independent if and only if $p_{Y|X}(y \mid x) = p_Y(y)$ for all possible values of x and y.

Exercise 2.4: Independence of Random Variables Show that the definition of independence of X and Y implies that

$$p_{X|Y}(x \mid y) = p_X(x)$$

and

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \qquad \text{for all possible values of } x \text{ and } y \qquad (2.7)$$

Given an arbitrary number N of random variables, they are said to be mutually independent if their joint pmf factors into the product of the corresponding N marginal pmf's.

Sometimes random variables may be independent but conditionally dependent; or, they may be dependent but conditionally independent. The definition of conditional independence is just what we expect: random variables X and Y are said to be conditionally independent given event A if and only if

$$P\{Y = y \mid X = x, (x, y) \in A\} = p_Y(y \mid A)$$

Exercise 2.5: Conditional Independence Show that for two random variables X and Y that are conditionally independent given event A,

$$p_{X,Y}(x, y \mid A) = p_X(x \mid A) p_Y(y \mid A) \qquad \text{for all } (x, y) \in A \tag{2.8}$$

Example 2: (continued)

In the minibus example, argue that N_F and N_M are not independent. Does there exist any nontrivial event A such that, given A, N_F and N_M are conditionally independent?

2.6 FUNCTIONS OF RANDOM VARIABLES

In Example 2 it was sometimes convenient to discuss the sum of the two random variables, reflecting the total number of people on the minibus. This is a particular case of defining one or more random variables as functions of other random variables. This function can be any rule for assigning an experimental value to the new random variable(s), given the experimental value(s) of the original random variable(s). Given original random variables X and Y, several possible functions of X and Y are as follows:

$$D = |X - Y|$$

$$R = \sqrt{X^2 + Y^2}$$

$$I = \begin{cases} 1 & \text{if } X + Y \text{ is an even number} \\ -1 & \text{otherwise} \end{cases}$$

$$V = \left\lfloor \frac{X}{Y} + 1 \right\rfloor \qquad \text{where } \lfloor x \rfloor = \text{greatest integer not exceeding } x$$

$$E = \text{Min}\,(X, Y)$$

The derivation of probability laws for functions of random variables will be a primary concern of Chapter 3.

2.7 EXPECTATION

Suppose that we have an experiment with random variable X and a function of X, $Y = g(X)$, which is itself a random variable. By this we mean that every experimental value x of the random variable X yields an experimental value $g(x)$ for the random variable $g(X) = Y$. Then the *expectation* or expected value of $g(X)$ is defined to be

$$E[g(X)] \equiv \sum_x g(x)p_X(x) = \overline{g(X)} \tag{2.9}$$

The conditional expected value of $g(X)$, given the experimental outcome is contained in event A, is

$$E[g(X) \mid A] = \sum_x g(x)p_X(x \mid A) = \overline{(g(X) \mid A)} \tag{2.10}$$

A key motivation for these definitions arises from large-sample theory, which reveals that if the experiment is performed independently many times, the empirically calculated *average* value of $g(\cdot)$ will probably be "very close to" $E[g(X)]$.[3] There are other motivations, too, such as z- and s-transforms, as we will see shortly.

Unfortunately, the word *expectation* or *expected value* of a random variable is perhaps one of the poorest word choices one encounters in probabilistic modeling. In practice, these words are often used interchangeably with *average* or *mean value* of a random variable. The problem here is that the mean or expected value of a random variable, when considered as a possible experimental value of the random variable, is usually quite unexpected and sometimes even impossible. For instance, a flip of a fair coin with "tails" yielding $X = 0$ and "heads" $X = 1$ results in an expected value $E[X] = \frac{1}{2}$, an impossible experimental outcome. Still, use of the term "expected value" persists and has caused considerable confusion in the minds of public administrators when reading consultants' reports or being briefed by unwary technical aides.

Two particular functions $g(X)$ will be of special interest in our work:

1. $g(X) = X$ yields the *mean value* or *expected value* of the random variable X,

$$E[X] = \bar{X} = \sum_x x p_X(x) \tag{2.11}$$

[3] For this statement to be true, $g(X)$ has to be "well behaved," where goodness of behavior usually implies that $E[g^2(X)]$ be finite.

2. $g(X) = (X - E[X])^2$ yields the *variance* or *second central moment* of the random variable X,

$$E[(X - E[X])^2] \equiv \sigma_X^2 = \sum_x (x - E[X])^2 p_X(x) \qquad (2.12)$$

Here σ_X, which is the square root of the variance, is the *standard deviation* of the random variable X.

Exercise 2.6: Expected Value of a Sum Show that the expected value of the sum of two arbitrary random variables X and Y is the sum of the two individual expected values (i.e., $E[X + Y] = E[X] + E[Y]$).

Exercise 2.7: Variance in Terms of Moments Show that

$$\sigma_X^2 = E[X^2] - E[X]^2.$$

Exercise 2.8: Variance of a Sum Show that for two *independent* random variables X and Y, the variance of the sum is the sum of the two individual variances (i.e., $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$).

Exercise 2.9: Expected Value of a Product Suppose that $X_1, X_2, \ldots, X_n$ are mutually independent random variables. Let

$$h(X_1, X_2, \ldots, X_n) = g_1(X_1)g_2(X_2) \ldots g_n(X_n)$$

Show that

$$E[h(X_1, X_2, \ldots, X_n)] = E[g_1(X_1)]E[g_2(X_2)] \ldots E[g_n(X_n)]$$

2.8 THE z-TRANSFORM

The expectation of one particular function $g(X)$, namely $g(X) = z^X$, is of wide use in computations and analysis for those discrete random variables X that take on only nonnegative integer experimental values. This expected value is defined to be the *z-transform* (or *discrete transform* or *geometric transform* or *moment-generating function*) of $p_X(x)$,

$$p_X^T(z) \equiv E[z^X] = \sum_{x=0}^{\infty} z^x p_X(x) \qquad (2.13)$$

Since $0 \leq p_X(x) \leq 1$, we are guaranteed that the summation identified in the definition of $p_X^T(z)$ will converge for $|z| < 1$. In fact, summations associated with z-transforms and their derivatives will usually converge for larger values of z, but restricting z so that $|z| < 1$ will assure us of no convergence problems in our work. Here we discuss briefly several useful properties of

z-transforms. The reader should be aware that material in subsequent chapters will *not* emphasize transform techniques and thus all of our coverage of transforms in this chapter may be considered to be optional material.

Given the z-transform of a pmf, we can uniquely recover the pmf. We do this by considering the definition of the z-transform,

$$p_X^T(z) = p_X(0) + p_X(1)z + p_X(2)z^2 + \dots + p_X(k)z^k + \dots \tag{2.14}$$

which yields through successive differentiation

$$p_X(k) = \frac{1}{k!}\left[\frac{d^k}{dz^k}p_X^T(z)\right]_{z=0} \qquad k = 0, 1, 2, \dots \tag{2.15}$$

The important moment-generating properties of the z-transform are obtained from the following relationships:

$$\left[\frac{dp_X^T(z)}{dz}\right]_{z=1} = \left[\sum_{x=0}^{\infty} xz^{x-1}p_X(x)\right]_{z=1} = E[X] \tag{2.16}$$

$$\left[\frac{d^2p_X^T(z)}{dz^2}\right]_{z=1} = \left[\sum_{x=0}^{\infty} x(x-1)z^{x-2}p_X(x)\right]_{z=1} = E[X^2] - E[X] \tag{2.17}$$

Manipulating (2.17), we can summarize the three key moment-generating properties of z-transforms as follows:

$$E[X] = \left[\frac{dp_X^T(z)}{dz}\right]_{z=1} \tag{2.18}$$

$$E[X^2] = \left[\frac{d^2p_X^T(z)}{dz^2}\right]_{z=1} + \left[\frac{dp_X^T(z)}{dz}\right]_{z=1} \tag{2.19}$$

$$\sigma_X^2 = \left[\frac{d^2p_X^T(z)}{dz^2}\right]_{z=1} + \left[\frac{dp_X^T(z)}{dz}\right]_{z=1} - \left(\left[\frac{dp_X^T(z)}{dz}\right]_{z=1}\right)^2 \tag{2.20}$$

Applications of these relationships are shown in the following section.

Exercise 2.10: z-Transform of a Sum Suppose that $X_1, X_2, \dots, X_n$ are mutually independent random variables. Let $S = X_1 + X_2 + \dots + X_n$. Show that

$$p_S^T(z) = p_{X_1}^T(z) \dots p_{X_n}^T(z)$$

2.9 OFTEN-USED PMF'S

Several specific pmf's arise frequently in practice, including in the analysis of urban service systems. In this section we identify and discuss some of these pmf's.

2.9.1 Bernoulli PMF

A random variable X whose probability law is a *Bernoulli pmf* can take on only two values, 0 and 1:

$$p_X(0) = P\{X = 0\} = 1 - p \qquad (2.21a)$$

$$p_X(1) = P\{X = 1\} = p \qquad (2.21b)$$

The mean and variance of a Bernoulli random variable are

$$E[X] = 0 \cdot (1 - p) + 1 \cdot p = p \qquad (2.22a)$$

$$\sigma_X^2 = E[X^2] - (E[X])^2 = 1^2 p - p^2 = p(1 - p) \qquad (2.22b)$$

The z-transform is $p_X^T(z) = (1 - p) + pz$. The Bernoulli pmf arises in simple trials having only two outcomes; it is also useful in the analysis of *set-indicator random variables* (see Section 3.3).

As an example of the use of a Bernoulli pmf, consider a police car performing "random" patrol. Each hour (on the hour) the patrol officer spins a wheel of fortune on the car's dashboard to see if the car should patrol a specially designated "high-crime zone" during the next hour. If police crime analysts have determined that approximately 25 percent of the time on random patrol should be spent in this zone, then an angle equal to $\pi/2$ on the wheel of fortune will be designated, "Patrol next hour in high-crime zone." Here the "patrol indicator" random variable X equals 1 if the car patrols the next hour in the high-crime zone; otherwise, $X = 0$. The expected value for X is $E[X] = p = 0.25$. The variance is $p(1 - p) = (0.25)(0.75) = 0.1875$. Each hour a separate (independent) experiment is performed. Thus, the total number of hours that the high-crime zone is patrolled during an 8-hour period is the sum of the individual X's corresponding to each hour. The idea of such a sequence of Bernoulli experiments will be developed further with the next two pmf's.

2.9.2 Geometric PMF

A random variable X has a *geometric pmf* if

$$p_X(x) = p(1 - p)^{x-1} \qquad x = 1, 2, \ldots \qquad (2.23)$$

The z-transform is

$$p_X^T(z) = \sum_{x=1}^{\infty} p(1 - p)^{x-1} z^x = \frac{pz}{1 - (1 - p)z} \qquad (2.24)$$

By differentiating $p_X^T(z)$ and substituting in (2.18) and (2.20), we obtain the mean and variance,

$$E[X] = \frac{1}{p} \tag{2.25}$$

$$\sigma_X^2 = \frac{1-p}{p^2} \tag{2.26}$$

One important interpretation of the geometric pmf involves the "first time until success" in a sequence of Bernoulli experiments (trials). Here "success" corresponds to the Bernoulli random value taking on the value 1. Suppose in the police example above that Y_i is the outcome of the Bernoulli trial conducted at the ith hour. Thus, if $Y_i = 1$, the high-crime zone is patrolled during the ith hour; otherwise, it is not patrolled that hour. Suppose that we (as observers) start looking at the high-crime zone during hour 1. We ask the question: Which hour X ($X = 1, 2, \ldots$) will be the first hour during which the high-crime zone will be patrolled? The probability that it will be patrolled during the first hour is simply p. The probability that it will be *first* patrolled during the second hour is $P\{Y_1 = 0, Y_1 = 1\}$, which by independence is $(1-p)p$. In general, the probability that it will be first patrolled during the kth hour is $P\{Y_1 = 0, Y_2 = 0, \ldots, Y_{k-1} = 0, Y_k = 1\}$, which by independence is $(1-p)^{k-1}p$. Thus, the random variable X is a geometrically distributed random variable which, when we substitute $p = 0.25$, has mean $E[X] = 1/0.25 = 4$ and variance $\sigma_X^2 = (\frac{3}{4})/(\frac{1}{4})^2 = \frac{3 \cdot 16}{4} = 12$ (and $\sigma_X = 2\sqrt{3} \approx 3.44$).

Question: What is the probability that the high-crime zone receives no patrol coverage during any particular 8-hour tour of duty?

Exercise 2.11: No Memory Property of Geometric PMF Suppose we have observed that the high-crime area has received no patrol coverage during the first k hours. Show that the probability law for the hour at which patrol first occurs, given this information, is the same as the original pmf, but shifted to the right k units. Thus, the geometric pmf has a *no-memory* property in the sense that the time (k hours) that we have invested waiting for the first hour of patrol coverage of the high-crime zone has not in any way reduced the mean or variance or any other measure of the *remaining time* we must wait until the first patrol.

2.9.3 Binomial PMF

A random variable W has a *binomial pmf* if

$$p_W(w) = \frac{n!}{w!\,(n-w)!}p^w(1-p)^{n-w} \qquad w = 0, 1, 2, \ldots, n \tag{2.27}$$

Here W can be interpreted to be the number of successes in n independent Bernoulli trials, each having success probability p. We can see this by writing

$$W = Y_1 + Y_2 + \ldots + Y_n$$

where Y_i is the ith Bernoulli random variable. Since the z-transform of the pmf of Y_i is $[(1 - p) + pz]$, we know from Exercise 2.10 that the z-transform of $p_W(\cdot)$ is

$$p_W^T(z) = [(1 - p) + pz]^n \tag{2.28}$$

Recalling the binomial theorem,

$$(a + b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i} \tag{2.29}$$

and the expanded form of the z-transform, (2.14), we obtain the binomial pmf shown in (2.27). By considering W to be the sum of n independent, identically distributed Bernoulli random variables, we obtain the mean and variance by inspection:

$$E[W] = np \tag{2.30a}$$

$$\sigma_W^2 = np(1 - p) \tag{2.30b}$$

In the police patrol example, with $p = 0.25$, the probability that the high-crime zone receives exactly w hours of patrol during an 8-hour tour of duty is

$$p_W(w) = \frac{8!}{w!\,(8 - w)!}\left(\frac{1}{4}\right)^w \left(\frac{3}{4}\right)^{8-w} \qquad w = 0, 1, \ldots, 8$$

2.9.4 Poisson PMF

A random variable K has a *Poisson pmf* if

$$p_K(k) = \frac{\mu^k e^{-\mu}}{k!} \qquad k = 0, 1, 2, \ldots;\ \mu > 0 \tag{2.31}$$

Substituting into the definition of the z-transform, we obtain

$$p_K^T(z) = e^{\mu(z-1)} \tag{2.32}$$

Differentiating (2.32) and substituting in (2.18) and (2.20), the mean and variance are found to be equal:

$$E[K] = \mu \tag{2.33a}$$

$$\sigma_K^2 = \mu \tag{2.33b}$$

In our work the Poisson pmf will arise most frequently in describing *Poisson processes*. These are processes in which "Poisson-type events" or "arrivals" are distributed totally randomly in time (or in space—see Section 3.4). In urban service systems, the Poisson process can be used as a reasonable model for the process generating fire alarms, police calls, ambulance calls, inquiries at a "little city hall," traffic passing through a lightly traveled intersection, breakdowns in a city's fleet of vehicles, letters arriving on the desk of a city administrator, the filled trash cans produced by a household, traffic violators on a given street segment, and so on. With the (time) Poisson process, we suppose that the process commences at $t = 0$ and that at random times $t_1, t_2, \ldots$, Poisson-type events occur (see Figure 2.6). Suppose that

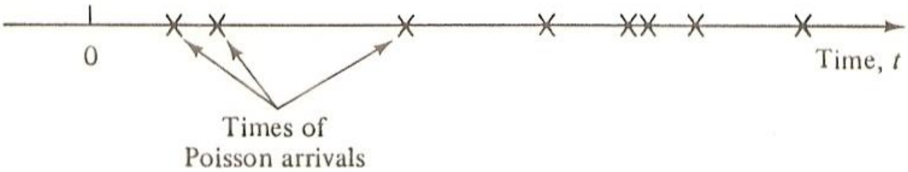


FIGURE 2.6 Poisson arrivals in time.

we are interested in the number of Poisson-type events $N(t)$ occurring in the time interval $[0, t]$. We prove in Section 2.12 that $N(t)$ has a Poisson pmf with mean λt:

$$P\{N(t) = k\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \qquad k = 0, 1, 2, \ldots \tag{2.34}$$

Here λ represents the average number of arrivals per unit time (see 2.12).

2.10 PROBABILITY DENSITY FUNCTIONS

Many random variables encountered in practice are distributed over a continuous rather than a discrete set of values. Examples include the time one waits at a bus stop until the next bus arrives, the tons of trash collected in a city on a given day, the distance a social worker in the field will travel on a given day, and the amount of electricity consumed by a household during a year. Just as probability mass functions (pmf's) allowed us to explore the probabilistic behavior of discrete random variables, probability density functions (pdf's) allow us to do the same for continuously distributed random variables.

We define a pdf for the (continuous) random variable X as follows:

$f_X(x)\,dx \equiv$ probability that the random variable X assumes an experimental value between x and $x + dx$ on a performance of the experiment

Note that our definition is not stated in terms of the probability that random variable X assumes exactly the value x; for a purely continuous random variable, this probability is zero. Thus, *in order to make any probability statement using pdf's, one must integrate the pdf* (even if only over an infinitesimal interval of length dx).

Some random variables occurring in practice are *mixed*; that is, they have a purely continuous part and they have a discrete part. An example could be the location of a bus at a random time along a straight-line street route; the bus might be viewed as uniformly distributed over the route except for a probability p_i of being located at $X = x_i$, the location of the ith stop ($i = 1, 2, \ldots, N$). In this case $\sum_{i=1}^{N} p_i$ is the "probability that the random variable X is discrete" and $(1 - \sum_{i=1}^{N} p_i)$ is the probability that it is continuous. To analyze the probabilistic behavior of X, we would treat separately each of the two components of X (discrete and continuous), and then combine the results using methods of conditional probability (see Problem 2.2). Thus, whenever possible throughout the remainder of this book, a continuous random variable is viewed as a purely continuous (rather than mixed) random variable. Still, on occasion it is necessary to consider a "continuous" random variable that has a positive probability of assuming a particular value. We do this with the *unit impulse function*, as shown later in this section.

Since probabilities must be nonnegative, we must have $f_X(x) \geq 0$. But unlike the pmf, whose value cannot exceed unity, there is no upper bound on the value of a pdf. A fundamental probabilistic statement involving a pdf relates the pdf to its cumulative distribution function (cdf),

$$F_X(x) \equiv P\{X \leq x\} = \int_{-\infty}^{x} f_X(y)\, dy \tag{2.35}$$

Since $P\{X \leq +\infty\} = 1$, we must have

$$\int_{-\infty}^{+\infty} f_X(y)\, dy = 1 \tag{2.36}$$

Example 3: Uniform PDF

Random variable U is distributed according to a uniform pdf (Figure 2.7) if

$$f_U(x) = \begin{cases} \dfrac{1}{b - a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \tag{2.37}$$

The cdf grows linearly from zero at $U = a$ to 1 at $U = b$ (Figure 2.8). The uniformly distributed random variable is often implied when the term "random" is used in problem statements, although we will attempt to avoid such ambiguous terminology here.

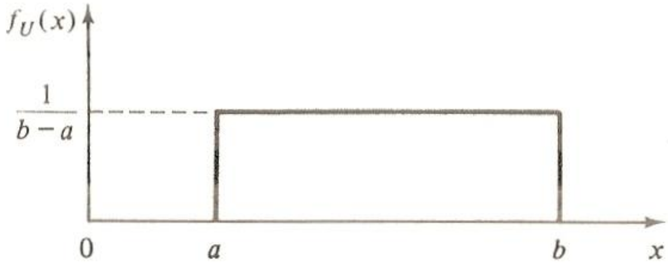


FIGURE 2.7 PDF of a uniformly distributed random variable.

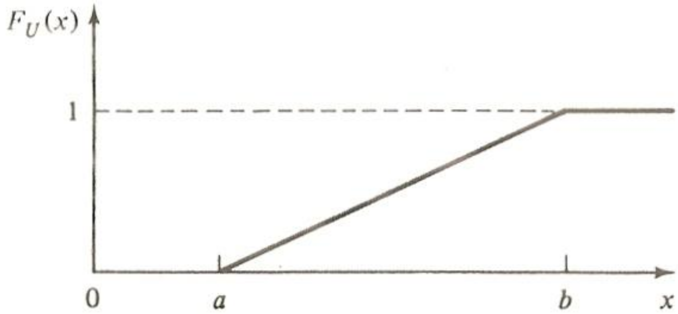


FIGURE 2.8 CDF of a uniformly distributed random variable.

The compound pdf allows us to study two or more continuous random variables simultaneously. For two random variables X and Y, their compound pdf is given by

$$f_{X,Y}(x, y)\, dx\, dy = P\{x \leq X \leq x + dx, y \leq Y \leq y + dy\} \qquad \text{all } x, y$$

In general, given N continuous random variables $X_1, X_2, \ldots, X_N$, there exists a corresponding N-argument pdf,

$$\begin{aligned} &f_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N)\, dx_1\, dx_2 \ldots dx_N \\ &\quad = P\{x_1 \leq X_1 \leq x_1 + dx_1, x_2 \leq X_2 \leq x_2 + dx_2, \ldots, x_N \leq X_N \leq x_N + dx_N\} \end{aligned}$$

Clearly,

$$f_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) \geq 0$$

$$\int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \ldots \int_{-\infty}^{+\infty} dx_N\, f_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) = 1$$

In cases involving multiple random variables $X_1, X_2, \ldots, X_N$, one may still be interested in the *marginal pdf* for $X_i, f_{X_i}(x_i)$, defined so that $f_{X_i}(x_i)\, dx_i \equiv P\{x_i \leq X_i \leq x_i + dx_i\}$. We can calculate the marginal from the joint pdf simply by integrating over all the other random variables:

$$\begin{aligned} f_{X_i}(x_i) &= \int dx_1 \int dx_2 \ldots \int dx_j \ldots \int dx_N \cdot \\ &\quad f_{X_1,X_2,\ldots,X_j,\ldots,X_N}(x_1, x_2, \ldots, x_j, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_N) \quad j \neq i \qquad (2.38) \end{aligned}$$

2.10.1 Conditional PDF's and Independence

When considering conditioning events and independence, the definitions from the discrete case carry over directly to the continuous case. For instance, given that an experimental outcome is contained in event A ($P\{A\} > 0$), the conditional compound pdf for two random variables X and Y is

$$f_{X,Y}(x, y \mid A) = \begin{cases} \dfrac{f_{X,Y}(x, y)}{P\{A\}} & (x, y) \in A,\ P\{A\} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.39}$$

If the event A is stated in terms of the specific experimental value of one of the random variables of the experiment, say $Y = 3.23$, we have a problem, because $P\{Y = 3.23\} = 0$. We circumvent this by considering an infinitesimal strip of width dy in the (X, Y) sample space and equate the conditioning event A to $\{y \leq Y \leq y + dy\}$. Then, employing the definition of conditional probability for the conditional event $\{x \leq X \leq x + dx\}$, given A, we have

$$f_{X|Y}(x \mid y)\, dx = \frac{f_{X,Y}(x, y)\, dx\, dy}{f_Y(y)\, dy}$$

Thus, the conditional pdf for one random variable, given the value of the other, is written

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \qquad f_Y(y) > 0 \tag{2.40a}$$

or, similarly,

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \qquad f_X(x) > 0 \tag{2.40b}$$

The general shape of this type of point-conditional pdf is determined by a vertical cut through the three-dimensional joint pdf at the fixed value of the conditioning random variable (Figure 2.9). The denominator, which is equal to the area of this cut, is simply a scaling factor.

Now, two continuous random variables X and Y are said to be *independent* if and only if

$$f_{Y|X}(y \mid x) = f_Y(y) \qquad \text{for all possible values of } x \text{ and } y$$

Exercise 2.12: Independence of Random Variables Show that the independence of X and Y implies that

$$f_{X|Y}(x \mid y) = f_X(x)$$

and

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \qquad \text{for all possible values of } x \text{ and } y$$

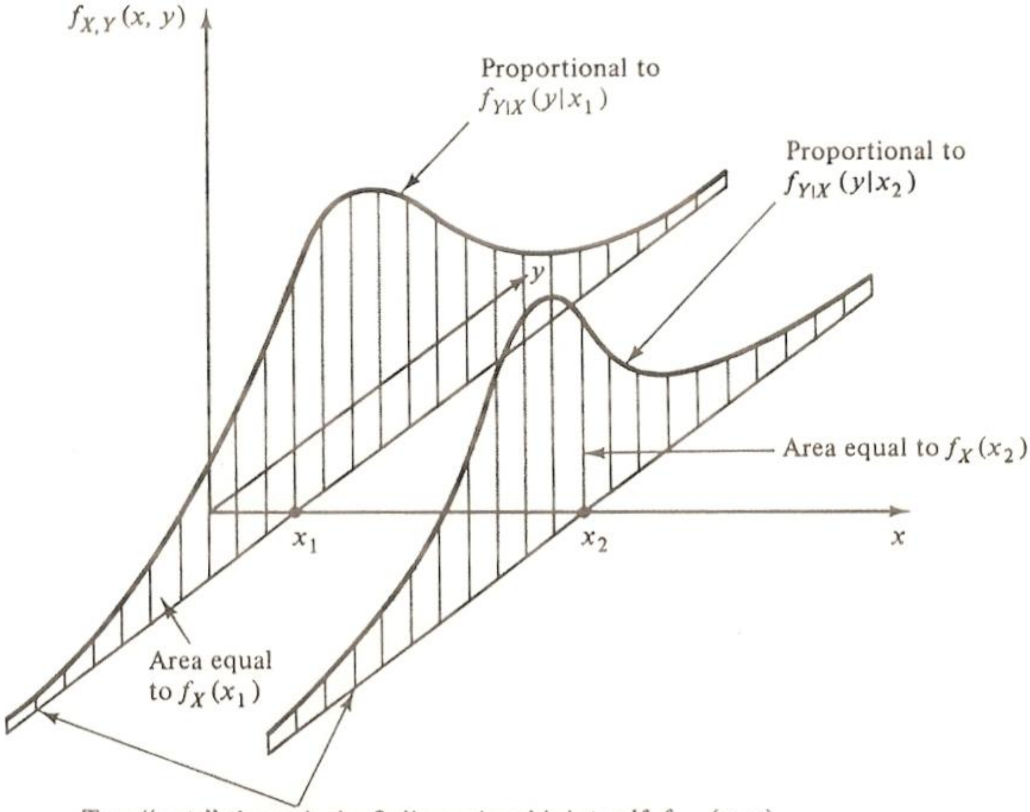


FIGURE 2.9 Pictorial depiction of $f_{Y|X}(y\,|\,x)$.

Given an arbitrary number N of continuous random variables, they are said to be mutually independent if their joint pdf factors into the product of the corresponding N marginal pdf's.

2.10.2 Expectation

Given a continuous random variable X with pdf $f_X(x)$, the expectation or expected value of the function $g(X)$ is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

All the results concerning expected values derived in Section 2.7 carry over in the obvious way, with summations replaced by integrations.

Exercise 2.13: Expected Values, Revisited Verify that the results of Exercises 2.6–2.9 also apply to continuously distributed random variables.

Example 1: (continued)

Here we continue our triangle problem initially described in Example 1, Section 2.1. We restate the problem as follows: "Two points X_1 and X_2 are marked randomly and independently on a stick of length 1 meter."

a. Determine the probability that a triangle can be formed with the three pieces obtained by cutting the stick at the marked points.

b. Determine the conditional pdf for X_1, given that a triangle can be formed.

c. Determine the conditional pdf for X_2, given that $X_1 = \frac{1}{4}$ and a triangle cannot be formed.

Solution:

a. First we must interpret the word "random." In the absence of any further information, the most reasonable interpretation is that X_1 and X_2 are *uniformly* and independently distributed over [0, 1]. Thus, the joint pdf for (X_1, X_2) is

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = \begin{cases} 1 & 0 \le x_1 \le 1, 0 \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Letting Δ be the event that a triangle can be formed, we recall that Δ corresponds to the two triangular regions of the (X_1, X_2) sample space shown in Figure 2.1. Since the area of each is $\frac{1}{8}$ and since the joint pdf is uniform with height 1, we obtain by inspection that $P\{\Delta\} = \frac{1}{4}$. If the pdf were not uniform, we would have to evaluate the following integral:

$$P\{\Delta\} = \int_{1/2}^{1} dx_1 \int_{x_1-1/2}^{1/2} dx_2 f_{X_1,X_2}(x_1, x_2) + \int_{0}^{1/2} dx_1 \int_{1/2}^{x_1+1/2} dx_2 f_{X_1,X_2}(x_1, x_2)$$

Since in this case $f_{X_1,X_2}(x_1, x_2) = 1$ over the regions of integration,

$$P\{\Delta\} = \int_{1/2}^{1} dx_1(1 - x_1) + \int_{0}^{1/2} dx_1\, x_1 = \frac{1}{4}$$

as we obtained by inspection.

b. Given that a triangle can be formed, the conditional (X_1, X_2) sample space comprises the two triangular regions over which we have just integrated. If one invokes the definition of the marginal pdf in terms of probabilities of lying within infinitesimal strips,

$$f_{X_1}(x_1 \mid \Delta)\, dx \equiv P\{x_1 \le X_1 \le x_1 + dx_1 \mid \Delta\}$$

then one can see from Figure 2.1 that this "strip probability" increases linearly from 0 to a maximum at $X_1 = \frac{1}{2}$ and then decreases linearly (and symmetrically) back to zero. Thus, by inspection,

$$f_{X_1}(x_1 \mid \Delta) = \begin{cases} cx & 0 \le x \le \frac{1}{2} \\ c(1-x) & \frac{1}{2} \le x \le 1 \end{cases}$$

where we find that $c = 4$ by the requirement that

$$\int_{-\infty}^{\infty} f_{X_1}(x_1 \mid \Delta)\, dx_1 = 1$$

Here again this problem is solvable by inspection since $f_{X_1,X_2}(\cdot, \cdot)$ is uniform over the region of interest. In general, we would obtain $f_{X_1}(x_1 \mid \Delta)$ by "integrating out" x_2:

$$f_{X_1}(x_1 \mid \Delta) = \begin{cases} \displaystyle\int_{1/2}^{(1/2)+x_1} f_{X_1,X_2}(x_1, x_2 \mid \Delta)\, dx_2 & 0 \le x_1 \le \frac{1}{2} \\ \displaystyle\int_{-(1/2)+x_1}^{1/2} f_{X_1,X_2}(x_1, x_2 \mid \Delta)\, dx_2 & \frac{1}{2} \le x_1 \le 1 \end{cases}$$

Here, since $f_{X_1,X_2}(x_1, x_2 \mid \Delta) = \text{constant} = 4$ over the region of integration, we have

$$f_{X_1}(x_1 \mid \Delta) = \begin{cases} 4(\frac{1}{2} + x_1 - \frac{1}{2}) = 4x_1 & 0 \le x_1 \le \frac{1}{2} \\ 4(\frac{1}{2} + \frac{1}{2} - x_1) = 4(1 - x_1) & \frac{1}{2} \le x_1 \le 1 \end{cases}$$

as anticipated.

c. The conditioning information is that $X_1 = \frac{1}{4}$ and a triangle cannot be formed. By inspection of Figure 2.1, the conditional pdf for X_2 is

$$f_{X_2|X_1}(x_2 \mid X_1 = \tfrac{1}{4}, \Delta') = \begin{cases} c' & 0 \le x_2 \le \frac{1}{2}, \frac{3}{4} \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

This pdf, which is displayed in Figure 2.10, is derived by considering a strip of infinitesimal width dx_1 at $x_1 = \frac{1}{4}$. Integration requires that $c' = \frac{4}{3}$.

We could continue the process of conditioning indefinitely and, in theory, incur no additional problems. For instance, let $A =$ event $|X_2 - \frac{1}{2}| > \frac{1}{8}$; this means that X_2 is either less than $\frac{3}{8}$ or greater than $\frac{5}{8}$. Then

$$f_{X_2|X_1}(x_2 \mid X_1 = \tfrac{1}{4}, \Delta', A) = \begin{cases} c'' & 0 \le x_2 \le \frac{3}{8},\ \frac{3}{4} \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

where $c'' = [\frac{3}{8} + (1 - \frac{3}{4})]^{-1} = \frac{8}{5}$. We could also find conditional moments, such as the conditional mean

$$E[X_2 \mid X_1 = \tfrac{1}{4}, \Delta', A] = \int_0^{3/8} x_2 \tfrac{8}{5}\, dx_2 + \int_{3/4}^{1} x_2 \tfrac{8}{5}\, dx_2 = \tfrac{37}{80}$$

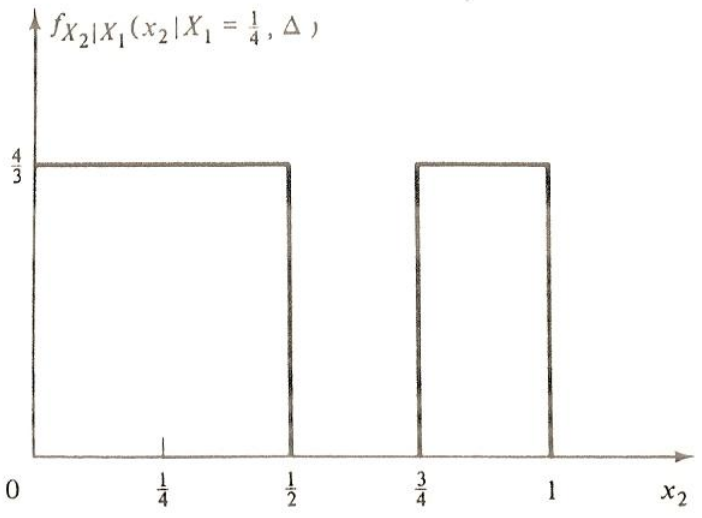


FIGURE 2.10 Conditional PDF for triangle problem.

or the conditional variance

$$\begin{aligned}\sigma^2_{\{X_2|X_1=1/4,\Delta',A\}} &= \int_0^{3/8} x_2^2 \tfrac{8}{5}\, dx_2 + \int_{3/4}^{1} x_2^2 \tfrac{8}{5}\, dx_2 - (\tfrac{37}{80})^2 \\ &= \tfrac{323}{960} - (\tfrac{37}{80})^2 \approx 0.12255\end{aligned}$$

2.10.3 Unit Impulse Function

On occasion we confront a "continuous" random variable that has a positive probability of assuming one or more particular experimental values. It is difficult to extend the notion of probability density functions to cover this case, since for any finite $f_X(x)$ the probability of assuming an experimental value in the interval x to $x + \Delta x$ is approximately $f_X(x)\,\Delta x$, becoming vanishingly small as $\Delta x \longrightarrow 0$.

A mechanism for circumventing this problem is to define a function $\mu_0(x)$ that has *area* equal to 1 at $x = 0$. That is,

$$\int_{-a}^{b} \mu_0(x)\, dx = 1 \tag{2.41}$$

for any positive a and b. Here $\mu_0(x)$, the *unit impulse function*, may be thought of as the limit as $\Delta x \longrightarrow 0$ of the function

$$g(x) = \begin{cases} \dfrac{1}{\Delta x} & 0 \leq x \leq \Delta x \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $g(x)$ is a legitimate pdf since it is always nonnegative and it integrates to 1. Since any probabilistic statement involving pdf's must always involve integration, we need not concern ourselves with the "value" of $\mu_0(x)$

$= \lim_{\Delta x \to 0} g(x)$ at $x = 0$, but only its integration properties near $x = 0$. Since $\mu_0(x)$ has area 1 at $x = 0$, the integral of any function $h(x)$ multiplied by $\mu_0(x)$ is simply $h(0)$, provided that the range of integration includes $x = 0$; that is,

$$\int_{-a}^{b} h(x)\mu_0(x)\,dx = h(0) \tag{2.42}$$

for any positive a and b.

For instance, a Bernoulli random variable having probability p of equaling 1 and $(1 - p)$ of equaling zero may be summarized by the following probability "density" function:

$$f_X(x) = (1 - p)\mu_0(x) + p\mu_0(x - 1)$$

The cdf behaves as we wish; for instance,

$$F_X(\tfrac{1}{2}) = \int_{-\infty}^{1/2} [(1 - p)\mu_0(x) + p\mu_0(x - 1)]\,dx = 1 - p$$

$$F_X(\tfrac{3}{2}) = \int_{-\infty}^{3/2} [(1 - p)\mu_0(x) + p\mu_0(x - 1)]\,dx = 1$$

In a similar manner, we could write any pmf as a pdf using impulse functions. We could also use impulse functions to depict the pdf of a mixed random variable.

2.10.4 The *s*-Transform

For continuous random variables the s-transform plays a role analogous to that played by the z-transform with discrete random variables. For a random variable X with pdf $f_X(x)$, the *s-transform* (or *exponential transform* or *Laplace transform*) is defined to be

$$f_X^T(s) \equiv E[e^{-sX}] = \int_{-\infty}^{+\infty} e^{-sx} f_X(x)\,dx \tag{2.43}$$

If s is considered to be a complex number, with real and imaginary parts, the integrals associated with s-transforms and their derivatives will be finite if the real part of s is zero. If $P\{X \leq a\} = 0$ for some finite a, these integrals will be finite as long as the real part of s is nonnegative. In the remainder of our work we will assume that the real part of s is chosen appropriately. And again, as with z-transforms, emphasis in subsequent chapters will not be given to transform techniques.

Given the s-transform of a pdf, one can uniquely recover the pdf. In general, this is done by contour integration in the complex plane. For those s-transforms whose numerators and demoninators factor into products of terms $(s - s_1)(s - s_2) \ldots$, the pdf can be recovered by partial fraction expansion.

By direct substitution into the definition of the s-transform, one can verify the following moment-generating properties:

$$E[X] = -\left[\frac{df_X^T(s)}{ds}\right]_{s=0} \tag{2.44}$$

$$E[X^2] = \left[\frac{d^2 f_X^T(s)}{ds^2}\right]_{s=0} \tag{2.45}$$

$$\sigma_X^2 = \left\{\frac{d^2 f_X^T(s)}{ds^2} - \left[\frac{df_X^T(s)}{ds}\right]^2\right\}_{s=0} \tag{2.46}$$

Applications of these relationships are shown in the following section.

Exercise 2.14: s-Transform of a Sum Suppose that $X_1, X_2, \ldots, X_n$ are mutually independent continuous random variables. Let $R = X_1 + X_2 + \ldots + X_n$. Show that

$$f_R^T(s) = f_{X_1}^T(s) f_{X_2}^T(s) \ldots f_{X_n}^T(s) \tag{2.47}$$

2.11 OFTEN-USED PDF'S

In this section we will introduce several pdf's that arise in the analysis of urban service systems. For future reference, we will tabulate the means and variances and other properties of interest.

2.11.1 Uniform PDF

In Section 2.10 we introduced the uniform pdf,

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

This pdf will arise often in the modeling of the distribution of entities in space (and in time, for that matter) and in the generating of random numbers for use in simulation experiments (see Chapter 7). By substituting directly into the definition, we find that

$$E[U] = \frac{a+b}{2} \tag{2.48a}$$

$$\sigma_U^2 = \frac{(b-a)^2}{12} \tag{2.48b}$$

$$f_U^T(s) = (e^{-as} - e^{-bs})[s(b-a)]^{-1} \tag{2.48c}$$

2.11.2 Exponential PDF

A random variable T is said to have an *exponential pdf* if

$$f_T(t) = \begin{cases} ae^{-at} & t \geq 0, \quad a > 0 \\ 0 & \text{otherwise} \end{cases}$$

The mean, variance, and s-transform are

$$E[T] = \frac{1}{a} \tag{2.49a}$$

$$\sigma_T^2 = \frac{1}{a^2} \tag{2.49b}$$

$$f_T^T(s) = \frac{a}{s+a} \tag{2.49c}$$

The exponential pdf arises in numerous contexts, including Poisson processes. It has a "no-memory" property similar to that of the geometric pmf. We demonstrate this in the following example.

Example 4: Jitney Rider

Suppose that a person walks to the side of the roadway to wait for a jitney, which will transport her to the next town. A *jitney* is a form of usually unscheduled transportation service involving minibuses or macro-taxicabs that travel back and forth between two towns or other centers. Jitney service, although uncommon in Europe and North America, is popular in many countries throughout the world. Suppose that, by analysis of past data or other means, it is known that the time required until the arrival of the next jitney is an exponentially distributed random variable with mean 10 minutes. Now suppose that our potential jitney rider has already waited 15.5 minutes and she wants to know the conditional mean additional time that she will have to wait.

Solution:

Let event $A = \{T > 15.5 \text{ minutes}\}$. Then we can find the conditional mean from the conditional pdf $f_T(t \mid A)$. But, by definition,

$$f_T(t \mid A) = \begin{cases} \dfrac{\frac{1}{10}e^{-t/10}}{P\{A\}} & \text{all } t \in A \\ 0 & \text{otherwise} \end{cases}$$

But "all $t \in A$" corresponds to $\{t > 15.5\}$ and $P\{A\} = 1 - F_T(15.5) = 1 - (1 - e^{-15.5/10}) = e^{-15.5/10}$. Thus,

$$f_T(t \mid A) = \begin{cases} \frac{1}{10}e^{-(t-15.5)/10} & t > 15.5 \\ 0 & \text{otherwise} \end{cases}$$

Hence, the conditional pdf is identical to the original exponential pdf but shifted to the right 15.5 units (Figure 2.11). By inspection, then, the conditional mean additional time that she has to wait remains 10 minutes. Here, "sunk investment" in waiting reaps no rewards in terms of reducing the remaining expected time until jitney arrival.

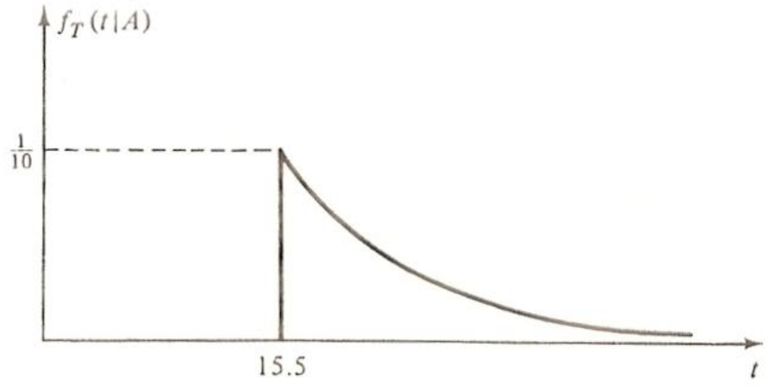


FIGURE 2.11 Exponential PDF shifted to the right.

Exercise 2.15: Jitney Rider, Revisited Redo Example 4 assuming that the pdf for the time until arrival of the next jitney is uniform between 2 and 18 minutes. How does sunk investment affect waiting time in this case?

2.11.3 Erlang PDF

A random variable L_k is said to be a kth-*order Erlang random variable* if its pdf is given by

$$f_{L_k}(x) = \begin{cases} \dfrac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} & k = 1, 2, \ldots; \quad x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.50}$$

The mean, variance, and s-transforms are

$$E[L_k] = \frac{k}{\lambda} \tag{2.51a}$$

$$\sigma^2_{L_k} = \frac{k}{\lambda^2} \tag{2.51b}$$

$$f^T_{L_k}(s) = \left(\frac{\lambda}{s+\lambda}\right)^k \tag{2.51c}$$

The Erlang pdf arises frequently in the application of Poisson processes. As we will show in Section 2.12, the kth-order Erlang pdf describes the probabilistic behavior of time until the kth arrival in a Poisson process. Note that for $k = 1$, the Erlang reduces to the familiar negative exponential pdf.

In fact, L_k may be thought of as the sum of k independent, identically distributed negative exponential random variables, each with mean $1/\lambda$; this provides a convenient way to understand intuitively the Erlang pdf and to remember the mean, variance, and s-transform.

2.11.4 Gaussian PDF

A random variable Y is said to have a *Gaussian* or *normal pdf* if

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-m_y)^2/2\sigma^2} \qquad -\infty < y < +\infty \tag{2.52}$$

The mean, variance, and s-transforms are

$$E[Y] = m_y \tag{2.53a}$$

$$\sigma_Y^2 = \sigma^2 \tag{2.53b}$$

$$f_Y^T(s) = e^{-m_y s + (1/2)\sigma^2 s^2} \tag{2.53c}$$

The Gaussian pdf arises most often in practice in applications of the *Central Limit Theorem*, which states (roughly) that the pdf of the sum of a large number of independent random variables approaches a Gaussian pdf with mean equal to the sum of the individual means and variance equal to the sum of the individual variances. The analyst of urban service systems should be familiar with this application of the Gaussian pdf. On occasion in this text we may invoke the Central Limit Theorem to approximate the pdf of a sum of random variables as a Gaussian random variable. Since we cannot obtain a closed-form expression for a partial integral of $f_Y(y)$, tables of the Gaussian pdf and cdf are widely available, for instance in mathematics and engineering handbooks.

2.12 POISSON PROCESS

We wish now to utilize our knowledge of pmf's, pdf's, and probabilistic modeling to analyze a very important process in urban services: the Poisson process. As mentioned earlier in Section 2.9, when introducing the Poisson pmf, the Poisson process is most often applied to occurrences of events in time. In an urban services context, these events could be requests for service, breakdowns of equipment, arrivals of vehicles at an intersection, or any of numerous other entities. So as not to confuse a Poisson event with events having an algebra (e.g., union and intersection), hereafter we will refer to Poisson-type events as *arrivals*, such as customers arriving at a queue. As we will see later, the concepts of a Poisson process can be extended to spatial applications in order to model, for instance, the locations of demands for service throughout a city.

First, we list the postulates of a Poisson process so that we can see the underlying physical assumptions necessary to give rise to the process. As we will see in Chapter 3, these postulates carry over in a natural way to spatial applications.

2.12.1 Postulates of a Poisson Process

There are four postulates associated with the Poisson process: First we state them informally, then mathematically:

1. *The probability that at least one Poisson arrival occurs in a small time period Δt is "approximately" $\lambda\, \Delta t$.* Here λ is called the arrival-rate parameter of the process. In applications, a numerical value for λ is found by measurement. Examples might be $\lambda = 10$ fire alarms per hour, or $\lambda = 62$ cars per hour passing through a tunnel, or $\lambda = 8.3$ unscheduled requests per day for a particular social service.
2. *The number of Poisson-type arrivals happening in any prespecified time interval of fixed length is not dependent on the "starting time" of the interval or on the total number of Poisson arrivals recorded prior to the interval.* For instance, if water-main breakdowns occur as a Poisson process, the number of breakdowns occurring in a particular day does not depend on the day being the tenth day of the month versus, say, the twentieth day of the month; nor does it depend on the number of breakdowns that occurred on the previous day or in the previous week.
3. *The numbers of arrivals happening in disjoint time intervals are mutually independent random variables.* Referring again to water-main breakdowns, say that we were interested in the number of breakdowns on September 28; this assumption would imply that knowledge (even partial knowledge) of the numbers of breakdowns on any days or combination of days other than September 28 would tell us nothing about the number of breakdowns on the 28th.
4. *Given that one Poisson arrival occurs at a particular time, the conditional probability that another occurs at exactly the same time is zero.* Thus, two or more arrivals cannot occur simultaneously. This may or may not be a good model for water-main breakdowns, but it certainly is not valid for the number of persons injured in auto accidents; given that an auto accident occurs at a particular time, it is an unfortunate fact that two or more persons may be injured at once.

These same four postulates can be stated mathematically more precisely:

1. The probability that at least one Poisson arrival occurs in a time period of duration τ is

$$p(\tau) = \lambda\tau + o(\tau)$$

where $o(\tau)$ is a generic expression for a term or collection of terms that "goes to zero faster than $k\tau$ as τ goes to zero" (for any constant k). Mathematically, $\lim_{\tau \to 0} o(\tau)/\tau = 0$. Note that for any finite sum of terms $o^N(\tau) = o_1(\tau) + o_2(\tau) + \cdots + o_N(\tau)$ such that $o_j(\tau)/\tau \longrightarrow 0$ as $\tau \longrightarrow 0$ ($j = 1, 2, \ldots, N$), we have $\lim_{\tau \to 0} o^N(\tau)/\tau = 0$.

2. Let $N(t)$ = total number of Poisson arrivals occurring in the interval $[0, t]$. We assume that $N(0) = 0$. For the interval $[t_1, t_2]$, the number of Poisson-type arrivals $[N(t_2) - N(t_1)]$ ($t_2 > t_1 \geq 0$) is dependent only on $(t_2 - t_1)$ and not on t_1 or $N(t_1)$.

3. If $0 \leq t_1 < t_2 \leq t_3 < t_4 \leq \ldots$, the numbers of arrivals occurring in disjoint time intervals $[N(t_2) - N(t_1)], [N(t_4) - N(t_3)], \ldots$ are mutually independent random variables.

4. The probability that two or more Poisson arrivals occur in a time interval of length τ is $o(\tau)$.

Given postulates 1–4, we now wish to prove the fundamental result for a Poisson process: that the number of Poisson arrivals occurring in a time interval of length t is Poisson-distributed with mean λt:

$$P\{N(t) = k\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \qquad k = 0, 1, 2, \ldots \tag{2.54}$$

Armed with this result, we can derive all the other interesting properties of a Poisson process.

Proof of (2.54): Let $P_m(t)$ denote the probability that exactly m arrivals occur in time t,

$$P_m(t) \equiv P\{N(t) = m\} \qquad m = 0, 1, 2, \ldots$$

We consider the time interval $[0, t]$ and expand it by an amount τ, as shown in Figure 2.12. We are interested in the probability that exactly $m + 1$ Poisson arrivals have occurred by time $t + \tau$. As shown in Figure 2.12, this event could occur with $m + 1$ arrivals occurring in $[0, t]$ and no arrivals in $[t, t + \tau]$, or m arrivals in $[0, t]$ and 1 arrival in $[t, t + \tau]$, and so on. Invoking independence of nonoverlapping intervals (postulate 3) and depen-

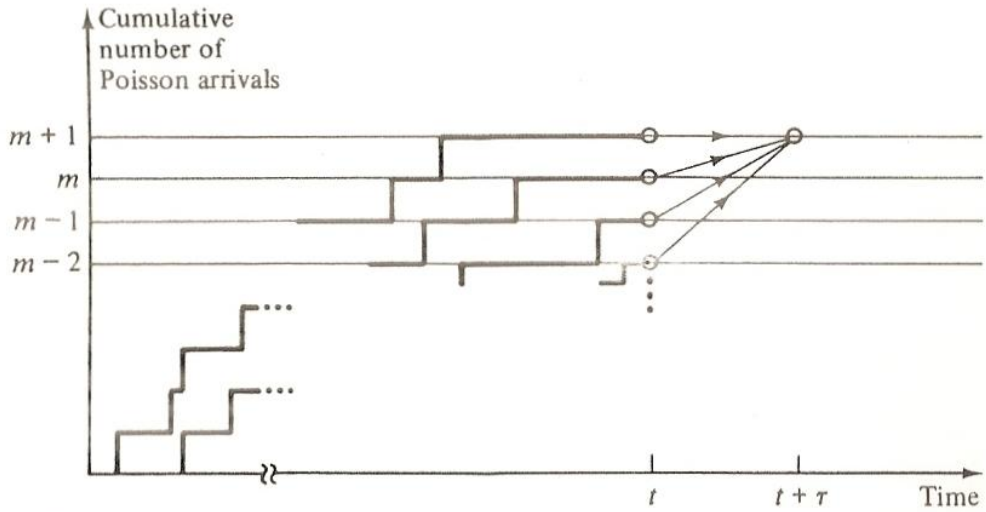


FIGURE 2.12 Possible system trajectories to reach $m + 1$ arrivals by time $t + \tau$.

dence only on time differences (postulate 2), we can write for $m = 1, 2, \ldots,$

$$P_{m+1}(t + \tau) = P_{m+1}(t)P_0(\tau) + P_m(t)P_1(\tau) + P_{m-1}(t)P_2(\tau) + \ldots$$

Letting τ become small, say $\tau \longrightarrow \Delta t$, we have, by postulates 1 and 4,

$$P_{m+1}(t + \Delta t) = P_{m+1}(t)[1 - \lambda\,\Delta t - o(\Delta t)] + P_m(t)[\lambda\,\Delta t + o\,(\Delta t)] + [\text{terms that are } o(\Delta t)]$$

Grouping the $o(\Delta t)$ terms together, we have

$$P_{m+1}(t + \Delta t) = P_{m+1}(t)(1 - \lambda\,\Delta t) + P_m(t)\lambda\,\Delta t + o(\Delta t)$$

or

$$\frac{P_{m+1}(t + \Delta t) - P_{m+1}(t)}{\Delta t} = -\lambda P_{m+1}(t) + \lambda P_m(t) + \frac{o(\Delta t)}{\Delta t}$$

Letting $\Delta t \longrightarrow 0$, which implies that $o(\Delta t)/\Delta t \longrightarrow 0$, we obtain the differential equation

$$\frac{dP_{m+1}(t)}{dt} = \lambda[P_m(t) - P_{m+1}(t)] \qquad m = 1, 2, \ldots \tag{2.55}$$

This equation makes sense intuitively: it states that the time rate of change of the probability of exactly $m + 1$ arrivals by time t is equal to the probability of exactly m arrivals in time t multiplied by the rate at which a "transition occurs" from m to $m + 1$ arrivals, *minus* the probability of already having $m + 1$ arrivals by t multiplied by the rate at which a transition occurs from $m + 1$ to $m + 2$ arrivals; the conditional transition rate in each case is λ. Similar logic can be used to develop sets of coupled differential equations for more complicated processes, say where λ is dependent on the number of previous arrivals or perhaps where arrivals can "depart," as in queueing systems. These ideas are expanded further in the discussion of "birth" processes in Chapter 3 and "birth-and-death" processes in Chapter 4.

While (2.55) holds for $m = 1, 2, \ldots$, we also require an equation for $m = 0$. In a manner similar to the derivation above, we obtain

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) \tag{2.56}$$

The intuitive interpretation here is directly analogous to that of (2.55). The solution to (2.56) is clearly

$$P_0(t) = ce^{-\lambda t} \qquad t \geq 0$$

with the constant c determined by the initial condition that $P_0(0) = 1$, implying that $c = 1$. Then, by substitution into (2.55), one proves by induction that

$$P_m(t) = \frac{(\lambda t)^m e^{-\lambda t}}{m!} \qquad m = 1, 2, \ldots$$

This completes the proof.

2.12.2 Interarrival Times

We are now interested in other properties of the Poisson process. For instance, suppose that we start observing the process at time $t = 0$ and we wish to know the pdf for the time of occurrence of the kth arrival. Define the random variable

$$L_k = \text{time of occurrence of the } k\text{th arrival} \qquad k = 1, 2, \ldots$$

Its pdf $f_{L_k}(\cdot)$ is sometimes called the "kth-order interarrival time distribution." We obtain $f_{L_k}(\cdot)$ simply by the following reasoning:

$$\begin{aligned} f_{L_k}(x)\,dx &\equiv P\{k\text{th arrival occurs in the interval } x \text{ to } x + dx\} \\ &= P\{\text{exactly } k-1 \text{ arrivals in the interval } [0, x] \textit{ and} \\ &\qquad \text{exactly one arrival in } [x, x+dx]\} \end{aligned}$$

By invoking the postulates of the Poisson process, plus (2.54), we can write

$$\begin{aligned} f_{L_k}(x)\,dx &= P\{\text{exactly } k-1 \text{ arrivals in } [0, x]\} \cdot \\ &\qquad\qquad P\{\text{exactly one arrival in } [x, x+dx]\} \\ &= \frac{(\lambda x)^{k-1} e^{-\lambda x}}{(k-1)!}[\lambda\,dx + o(dx)] \end{aligned}$$

Since dx is infinitesimal, we can ignore the $o(dx)$ term, and obtain the kth-order interarrival time distribution

$$f_{L_k}(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \qquad x \geq 0; \quad k = 1, 2, \ldots \tag{2.57}$$

We have the fundamental result that the kth-order interarrival time distribution for a Poisson process is a kth-order Erlang pdf.

Setting $k = 1$ in (2.57), we have the first-order arrival distribution

$$f_{L_1}(x) = \lambda e^{-\lambda x} \qquad x \geq 0 \tag{2.58}$$

This is our familiar exponential pdf with its now famous "no-memory" property. This result proves that the Poisson process is a no-memory process, that is, that future arrivals do not depend on the number or times of occurrences of previous arrivals.

2.12.3 Unordered Arrival Times

We now know that arrivals occur singly in a Poisson process with interarrival times between successive arrivals distributed as negative exponential random variables. But suppose that we turn things around a bit and say that we have waited through the interval $[0, t]$ and that exactly m Poisson arrivals have occurred in $[0, t]$. And, rather than queue up in some orderly fashion (like first arrivals first in line), all the arrivals are mixed together, such as the students sitting in a classroom or the patients sitting in the waiting room of an outpatient clinic. We wish to derive a property of the Poisson process which in part is responsible for its nickname "most random of random processes."

We call the m arrivals the *unordered* arrivals of a Poisson process, since they are not ordered in accordance with the time of their arrival. We wish to determine the probabilistic behavior of the m unordered arrival times. To do this, we partition the interval $[0, t]$ into $2m + 1$ arbitrary subintervals a_i and b_i, as shown in Figure 2.13. Suppose that we are interested in the event E_m that exactly one arrival occurred in each of the subintervals b_i and that no arrival occurred in any of the subintervals a_i. We wish to calculate the probability of E_m occurring, given that exactly m arrivals occurred in $[0, t]$.

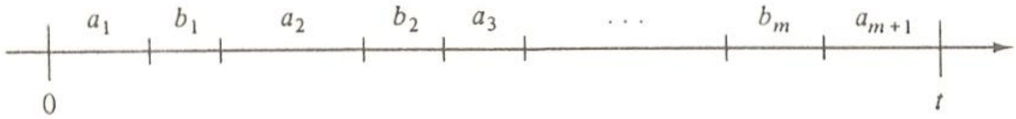


FIGURE 2.13 Partitioning of $[0, t]$ for unordered arrival times.

(Clearly, the probability of E_m is zero for any numbers of arrivals other than m.) Thus, invoking the definition of conditional probability, we want

$$\begin{aligned} P\{E_m \mid m \text{ Poisson arrivals in } [0, t]\} &= \frac{P\{E_m \text{ and } m \text{ Poisson arrivals in } [0, t]\}}{P\{m \text{ Poisson arrivals in } [0, t]\}} \\ &= \frac{P\{m \text{ Poisson arrivals in } [0, t] \mid E_m\} \cdot P\{E_m\}}{[(\lambda t)^m / m!] e^{-\lambda t}} \\ &= \frac{P\{E_m\}}{[(\lambda t)^m / m!] e^{-\lambda t}} \end{aligned} \tag{2.59}$$

But E_m is the union of $2m + 1$ events that, by the Poisson postulates, are mutually independent. For m of the events, we want the probability of exactly one arrival in a subinterval b_k, and this is simply

$$P\{\text{exactly one arrival in } b_k\} = \frac{\lambda b_k e^{-\lambda b_k}}{1!} \qquad k = 1, 2, \ldots, m$$

For $m + 1$ of the events, we want the probability of no arrivals in a subinterval a_k and this is

$$P\{\text{zero arrivals in } a_k\} = e^{-\lambda a_k} \qquad k = 1, 2, \ldots, m + 1$$

Invoking mutual independence to multiply the individual probabilities, we have, upon substitution into (2.59),

$$P\{E_m \mid m \text{ Poisson arrivals in } [0, t]\} = \frac{(\lambda b_1 \lambda b_2 \ldots \lambda b_m e^{-\lambda b_1} e^{-\lambda b_2} \ldots e^{-\lambda b_m})(e^{-\lambda a_1} e^{-\lambda a_2} \ldots e^{-\lambda a_{m+1}})}{[(\lambda t)^m / m!] e^{-\lambda t}}$$

Since $\sum b_k + \sum a_k = t$, the exponentials divide out and we have

$$P\{E_m \mid m \text{ Poisson arrivals in } [0, t]\} = \frac{b_1 b_2 \ldots b_m}{t^m} m! \tag{2.60}$$

But suppose that the process had been one in which we took each potential arrival, say a person, and "tossed" him/her "at random" into the interval $[0, t]$, with successive tosses being independent. The probability that any particular arrival "ends up" in interval b_k is then simply b_k/t, since the arrival time of such a person is uniformly distributed over $[0, t]$. The probability α that the m arrivals distributed over $[0, t]$ in this way each end up as the single occupant of one subinterval b_k is then simply

$$\alpha = \left(\frac{b_1}{t}\right)\left(\frac{b_2}{t}\right)\left(\frac{b_3}{t}\right) \cdots \left(\frac{b_m}{t}\right) m! \tag{2.61}$$

where the factorial term arises from the number of distinct ways in which the arrivals could be situated in the subintervals; that is, there are $m!$ points in the experiment's sample space having the property we desire, and each point has the same probability.

Since (2.60) and (2.61) are identical, we thus have derived the following result: *the unordered arrival times in a Poisson process are independently, uniformly distributed over the fixed time interval of interest.*

2.12.4 Multiple Independent Poisson Processes

Suppose that there are two Poisson processes operating independently, with arrival rates λ_1 and λ_2, respectively. $N_1(t)$ and $N_2(t)$ are the respective cumulative numbers of arrivals through time t. Then the *combined* or *pooled* process has a cumulative number of arrivals equal to $N(t) = N_1(t) + N_2(t)$. A fundamental property of independent Poisson processes is that their pooled process is also a Poisson process with arrival-rate parameter equal to the sum of the individual arrival rates. Thus, $N(t)$ has a Poisson distribution with mean $(\lambda_1 + \lambda_2)t$. This result extends in the obvious way to more than two independent Poisson processes. There are many ways to prove this result, but the simplest is just to observe that the pooled process satisfies each of the four postulates of the Poisson process.

We are often confronted with the following question: Given two independently operating Poisson processes with rate parameters λ_1 and λ_2, respectively, what is the probability that an arrival from process 1 (a "type 1" arrival) occurs *before* an arrival from process 2 (a "type 2" arrival)? To solve this problem, let the two arrival times of interest be denoted by X_1 and X_2, for processes 1 and 2, respectively. We want to compute $P\{X_1 < X_2\}$. Invoking our knowledge of Poisson processes, we know that the pdf's for X_1 and X_2 are negative exponentials with means λ_1^{-1} and λ_2^{-1}, respectively. Thus, because of independence, the joint pdf is

$$f_{X_1,X_2}(x_1, x_2) = \lambda_1\lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2} \qquad \begin{matrix} x_1 \geq 0 \\ x_2 \geq 0 \end{matrix}$$

Integrating over the part of the positive quadrant for which $x_1 < x_2$, we have

$$\begin{aligned} P\{X_1 < X_2\} &= \int_0^\infty dx_1 \int_{x_1}^\infty dx_2 \lambda_1 \lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2} \\ &= \int_0^\infty dx_1\, \lambda_1 e^{-\lambda_1 x_1}(e^{-\lambda_2 x_1}) \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2}\int_0^\infty e^{-u}\, du \end{aligned}$$

or, since the integral equals 1,

$$P\{X_1 < X_2\} = \frac{\lambda_1}{\lambda_1 + \lambda_2} \tag{2.62}$$

This result makes sense intuitively: The probability that a type 1 arrival occurs before a type 2 arrival is equal to the fraction of the pooled arrival rate comprising type 1 arrivals.

This important result can be derived in a number of other ways, as well. Suppose that we are awaiting the first Poisson-type arrival and finally we are given the conditional information that the arrival occurred in the interval $[t, t + dt]$. Let

$P\{\text{type 1} \mid \text{one arrival in } [t, t + dt]\}$ = probability that the arrival is type 1 given that it occurs in $[t, t + dt]$

Invoking conditional probability,

$$P\{\text{type 1} \mid \text{one arrival in } [t, t + dt]\} = \frac{P\{\text{type 1 } \textit{and} \text{ it occurs in } [t, t + dt]}{P\{\text{one arrival occurs in } [t, t + dt]\}}$$

$$= \frac{\lambda_1\, dt + o(dt)}{(\lambda_1 + \lambda_2)\, dt + o(dt)}$$

Since dt is infinitesimal, we can ignore the $o(dt)$ term, thereby obtaining $\lambda_1/(\lambda_1 + \lambda_2)$ for the conditional probability, conditioned on the infinitesimal arrival interval. We uncondition the probability by integrating over all possible arrival intervals, weighted by the probability of their occurrence,

$$P\{\text{type 1}\} = \int_0^\infty P\{\text{type 1} \mid \text{one arrival in } [t, t + dt]\}(\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)t}\, dt$$

$$= \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Yet a third way of deriving the result involves consideration of a long time period of length T. During that period the expected total number of arrivals is $(\lambda_1 + \lambda_2)T$; the expected number of type 1 arrivals is $\lambda_1 T$. Thus, roughly speaking, the fraction of total arrivals that is type 1 is

$$\frac{\lambda_1 T}{(\lambda_1 + \lambda_2)T} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

a result in agreement with the more formally derived (2.62).

A final point about this result: in examining a Poisson process we occasionally only have to look at the arrival instants, distinguishing between, say, type 1 and type 2 arrivals, not caring about the exact times of their occurrence. Then, because of the no-memory property of Poisson processes, each arrival instant becomes an independent Bernoulli experiment, with probability of a type 1 arrival equal to $\lambda_1/(\lambda_1 + \lambda_2)$ and of a type 2 equal to $\lambda_2/(\lambda_1 + \lambda_2)$. Thus, when considering questions of the types of Poisson arrivals occurring (and not their times), we may invoke ideas of Bernoulli experiments which often lead to geometric distributions, binomial distributions, and other pmf's associated with independent *discrete* random variables.

We apply most of our results for a Poisson process to a "pedestrian crossing" problem at the end of the chapter.

2.13 RANDOM INCIDENCE

The Poisson process is one of many stochastic processes that one encounters in urban service systems. The Poisson process is one example of a "point process" in which discrete events (arrivals) occur at particular points in time. For a general point process having its zeroth arrival at time T_0 and the remaining arrivals at times $T_1, T_2, T_3, \ldots$, the interarrival times are

$$Y_1 = T_1 - T_0$$
$$Y_2 = T_2 - T_1$$
$$\vdots$$
$$Y_k = T_k - T_{k-1}$$

Such a stochastic process is fully characterized by the *family* of joint pdf's $f_{T_0,Y_{n_1},Y_{n_2},\ldots,Y_{n_p}}(t_0, y_{n_1}, y_{n_2}, \ldots, y_{n_p})$ for all integer values of p and all possible combinations of different $n_1, n_2, \ldots$, where each n_i is a positive integer denoting a particular interarrival time. Maintaining the depiction of a stochastic process at such a general level, although fine in theory, yields an intractable model and one for which the data (to estimate all the joint pdf's) are virtually impossible to obtain. So, in the study of stochastic processes, one is motivated to make assumptions about this family of pdf's that (1) are realistic for an important class of problems and (2) yield a tractable model.

We wish to consider here the class of point stochastic processes for which the marginal pdf's for all of the interarrival times (Y_k) are identical. That is, we assume that

$$f_Y(x) = f_{Y_1}(x) = f_{Y_2}(x) = \ldots = f_{Y_k}(x) = \ldots$$

Thus, for Y_k, if we selected any one of the family of joint pdf's $f_{Y_{n_1},Y_{n_2},\ldots,Y_{n_p}}(y_{n_1}, y_{n_2}, \ldots, y_k, \ldots, y_{n_p})$ and "integrated out" all variables except y_k, we would obtain $f_Y(\cdot)$. Note that we have said nothing about independence of the Y_k's.

They need not be mutually independent, pairwise independent, or conditionally independent in any way. For the special case in which the Y_k's *are* mutually independent, the point process is called a *renewal process.* The Poisson process is a special case of a renewal process, being the only continuous-time renewal process having "no memory." However, the kind of process we are considering can exhibit both memory and dependence among the inter-event times. In fact, the dependence could be so strong that once we know the value of one of the Y_k's we might know a great deal (perhaps even the exact values) of any number of the remaining Y_k's.

Example 5: Bus Stop

Consider a potential bus passenger arriving at a bus stop. The kth bus arrives Y_k time units after the $(k - 1)$st bus. Here the Y_k's are called bus *headways.* The probabilistic behavior of the Y_k's will determine the probability law for the waiting time of the potential passenger (until the next bus arrives). Here it is reasonable to assume that the Y_k's are identically distributed but not independent (due to interactions between successive buses). One could estimate the pdf $f_Y(\cdot)$ simply by gathering data describing bus interarrival times and displaying the data in the form of a histogram. (This same model applies to subways and even elevators in a multielevator building.)

Example 6: Police Patrol

Consider the process of passings of a police patrol car by a residence or business. Successive passings may be identically distributed but not independent. For instance, the patrolling officer may think to himself or herself: "Well, I've passed that address three times during the last two hours, so I won't go by there again until tomorrow."

In situations such as these, for which we know $f_Y(\cdot)$ or at least the mean and variance of Y, we are often interested in the following problem. An individual, say a potential bus passenger or a homeowner looking for a police patrol car, starts observing the process at a *random time*, and he or she wishes to obtain the probability law (or at least the mean) of the time he or she must wait until the *next* arrival occurs. In various applications this time could be the waiting time for a bus, subway, or elevator or the time until arrival of a patrol car. This is said to be a problem of *random incidence*, since the individual observer is incident to the process at a random time. The random time assumption is important: the time of random incidence of the observer can in no way depend on the past history of actual arrival times in the process.

We now derive the probability law for V, the time from the moment of random incidence until the next arrival occurs. We do this for continuous random variables since the same reasoning applies in the discrete case. The derivation proceeds in stages, first conditioning on W, the length of the inter-arrival *gap* entered by random incidence. For instance, the gap in which a potential bus passenger arrives has length equal to the sum of two time intervals: (1) the time between the arrival of the most recent bus and the arrival of the potential passenger, and (2) the time between the passenger's arrival and the arrival of the next bus. We now argue that the probability that the gap entered by random incidence assumes a value between w and $w + dw$ is proportional to *both* the relative frequency of occurrence of such gaps $f_Y(w)\,dw$ *and* the duration of the gap w. That is,

$$f_W(w)\,dw \propto w f_Y(w)\,dw$$

Since a pdf must integrate to 1, the constant of proportionality must be $1/E[Y]$, so that

$$f_W(w) = \frac{w f_Y(w)}{E[Y]} \tag{2.63}$$

This result says that random incidence favors gaps of longer duration in direct proportion to their duration. The argument for this result is most simply given by example. Given two gap lengths w_1 and $w_2 = 2w_1$ for which the relative frequencies are identical $[f_Y(w_1)\,dw = f_Y(w_2)\,dw]$, then one is twice as likely to enter the gap of length $2w_1$ compared to the gap of length w_1. Or, given the same two gap lengths, w_1 and $w_2 = 2w_1$, for which the relative frequency of the large gap length is only half that of the smaller $[f_Y(w_2)\,dw = \frac{1}{2}f_Y(w_1)\,dw]$, we are equally likely to enter either of the two types of gaps; here the doubling of relative frequency for w_1 "makes up for" the doubling of duration of w_2.

Now, given that we have entered a gap of length w by random incidence, we are *equally likely* to be anywhere within the gap. More precisely, there is a constant probability of being in any interval τ to $\tau + h$ for any fixed $h > 0$, assuming that $[\tau, \tau + h]$ is fully contained within the gap. Thus, given w, the time until gap completion (i.e., the time until the next event) has a uniform pdf:

$$f_{V|W}(v \mid w) = \frac{1}{w} \qquad 0 \le v \le w \tag{2.64}$$

Combining our two results so far, we obtain the joint pdf for V and W:

$$\begin{aligned} f_{V,W}(v, w) &= f_{V|W}(v \mid w) f_W(W) \\ &= \frac{1}{w}\frac{w f_Y(w)}{E[Y]} \qquad 0 \le v \le w \le \infty \end{aligned}$$

The marginal for V, which is what we want, is formed simply by "integrating out" w,

$$f_V(v) = \int_v^{\infty} \frac{f_Y(w)}{E[Y]}\, dw$$

which can be expressed in terms of the cdf for Y,

$$f_V(v) = \frac{1 - F_Y(v)}{E[Y]} \qquad v \geq 0 \tag{2.65}$$

Question: Assuming that $F_Y(0) = 0$, does this result make intuitive sense for values of v near zero?

Example 5: (continued)

Suppose that buses maintain perfect headway; that is, they are always T_0 minutes apart. Then

$$F_Y(v) = \begin{cases} 0 & v < T_0 \\ 1 & v \geq T_0 \end{cases}$$

$$f_V(v) = \begin{cases} \dfrac{1}{T_0} & 0 \leq v \leq T_0 \\ 0 & \text{otherwise} \end{cases}$$

That is, the time until the next bus arrives, given random incidence, is uniformly distributed between 0 and T_0, with a mean $E[V] = T_0/2$, as we might expect intuitively.

Example 6: (continued)

Suppose that police cars patrol in a completely random manner, with car passings occurring according to a Poisson process with mean rate λ passings per day. Then interpassing times are distributed as negative exponential random variables with mean $1/\lambda$. Hence,

$$F_Y(v) = 1 - e^{-\lambda v} \qquad v \geq 0$$

and

$$f_V(v) = \frac{1 - (1 - e^{-\lambda v})}{1/\lambda}$$

$$= \lambda e^{-\lambda v} \qquad v \geq 0$$

as we expect from the no-memory property of Poisson processes.

Example 7: Clumped Buses

Suppose that buses along a particular route are on schedule "half the time" and "clumped" together in pairs the other half of the time. That is, for 50 percent of the day (which 50 percent is unpredictable), the bus headways are exactly T_0. For the remaining 50 percent of the day, (crowded) buses arrive in pairs, the time between each pair being $2T_0$. Note that in this case, given paired buses, half of the bus interarrival times are $2T_0$ and half are 0 (!), because of zero elapsed time between two paired buses. Thus,

$$F_Y(v) = \begin{cases} \frac{1}{4} & 0 \le v < T_0 \\ \frac{3}{4} & T_0 \le v < 2T_0 \\ 1 & 2T_0 \le v \end{cases}$$

$$E[Y] = 0 \cdot \tfrac{1}{4} + T_0 \cdot \tfrac{1}{2} + 2T_0 \cdot \tfrac{1}{4} = T_0$$

The pdf for bus waiting time, given random incidence, becomes

$$f_V(v) = \begin{cases} \dfrac{3}{4T_0} & 0 \le v < T_0 \\ \dfrac{1}{4T_0} & T_0 \le v < 2T_0 \\ 0 & \text{otherwise} \end{cases}$$

The mean wait for a bus is

$$E[V] = \int_0^\infty v f_V(v)\, dv = \tfrac{3}{4} T_0$$

which represents a 50 percent increase over the case of perfect headways. (Note that successive bus interarrival times are not independent in this case.)

Sometimes we may only be interested in $E[V]$, the mean time from the moment of random incidence until the next arrival occurs. Or, we may have only partial statistics about Y, perhaps the first few moments, and we desire to obtain at least $E[V]$. We may compute $E[V]$ simply by conditioning on the length of the gap entered and integrating over all possible gap lengths:

$$E[V] = \int_0^\infty E[V \mid w] f_W(w)\, dw$$

But, given $W = w$, V is uniformly distributed between 0 and w [see (2.64)]; thus,

$$E[V \mid w] = \tfrac{1}{2} w$$

Using (2.63), we can now write

$$E[V] = \int_0^\infty \frac{1}{2} w \frac{w f_Y(w)}{E[Y]}\, dw$$

$$E[V] = \frac{E[Y^2]}{2E[Y]} = \frac{\sigma_Y^2 + E^2[Y]}{2E[Y]} \qquad (2.66)$$

Thus, the *mean time from random incidence until the next event depends only on the mean and variance of the inter-event time Y.*

Example 5: (continued)

For buses with perfect headways, $E[Y] = T_0$ and $\sigma_Y^2 = 0$. Substituting in (2.66), we have

$$E[V] = \frac{0 + T_0^2}{2T_0} = \frac{T_0}{2}$$

as previously computed.

Example 6: (continued)

For perfectly random police patrol,

$$E[Y] = \frac{1}{\lambda} \qquad \sigma_Y^2 = \frac{1}{\lambda^2}$$

and thus

$$E[V] = \frac{1/\lambda^2 + 1/\lambda^2}{2(1/\lambda)} = \frac{1}{\lambda}$$

as expected.

Example 7: (continued)

For buses that are clumped in pairs 50 percent of the time,

$$E[Y] = T_0$$

$$E[Y^2] = 0 \cdot \tfrac{1}{4} + T_0^2 \cdot \tfrac{1}{2} + (2T_0)^2\, \tfrac{1}{4} = \tfrac{3}{2}\, T_0^2$$

Substituting in (2.66), we obtain the earlier result

$$E[V] = \frac{\frac{3}{2} T_0^2}{2T_0} = \tfrac{3}{4}\, T_0$$

It is interesting to examine (2.66) to acquire an understanding of the range of plausible values of $E[V]$. The minimum possible mean waiting time until the next arrival is $\frac{1}{2}E[Y]$, which occurs for a "perfectly scheduled" system (i.e., $\sigma_Y^2 = 0$). For systems with temporal irregularity comparable to that of the Poisson process, namely those for which the standard deviation of interarrival times σ_Y equals the mean $E[Y]$, the waiting time $E[V]$ is equal to the mean $E[Y]$. Thus, for such systems one incurs the same average wait arriving at random as an observer arriving immediately after the most recent event (e.g., bus). Intuitively, in this case one half the mean wait is due to the average spacing between successive buses and the other half is due to uncertainties (i.e., randomness) in the arrival process. For systems with irregularity greater than the Poisson process, namely those for which $\sigma_y > E[Y]$, the mean wait of the observer can assume any value greater than $E[Y]$; here, somewhat surprisingly, one waits longer (on the average) arriving at a random time compared to arriving just after the most recent event. An example of such a situation would be arriving at the turnstiles of a sports stadium, say Yankee Stadium in New York, and waiting for the next sports fan to pass through the turnstile. The wait is likely to be very small (say in the order of seconds) if one arrives just after the arrival of a random fan (thereby guaranteeing that a baseball game is about to be played) compared to arriving at some random time during the year (in which case the wait is likely to be quite long if one arrives, say, in December).

2.14 PEDESTRIAN CROSSING PROBLEM

In this section we study a problem involving the timing of a pedestrian-street-crossing light. The problem provides a good illustration of many of the probabilistic modeling techniques that we have reviewed in this chapter, particularly the Poisson process. It also provides an example of applying mathematical modeling to the evaluation of technological alternatives in an urban setting.

The problem is as follows. Pedestrians approach from the $\begin{Bmatrix}\text{left}\\ \text{right}\end{Bmatrix}$ side of the crossing in a Poisson manner with average arrival rate $\begin{Bmatrix}\lambda_L\\ \lambda_R\end{Bmatrix}$ arrivals per minute (Figure 2.14). Each pedestrian then waits until a light is flashed, at which time all waiting pedestrians must cross. We refer to each time the light is flashed as a "dump" and assume that a dump takes zero time (i.e., pedestrians cross instantly). Assume that the left and right arrival processes are independent.

We wish to analyze three possible decision rules for operating the light:

Rule A: Dump every T minutes.

Rule B: Dump whenever the *total* number of waiting pedestrians equals N_0.

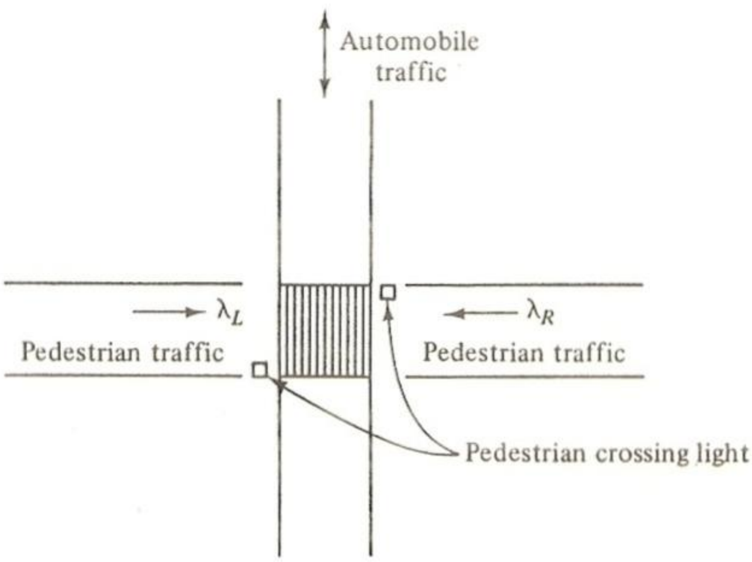


FIGURE 2.14 Pedestrian crossing problem.

Rule C: Dump whenever the *first* pedestrian to arrive after the previous dump has waited T_0 minutes.

Presumably, implementation of each rule requires a particular type of technology with its attendant costs, and thus it is important to determine the operating characteristics of each in order to understand tradeoffs between performance and cost.

For each decision rule, determine:

1. The expected number of pedestrians crossing *left to right* on any dump.
2. The probability that zero pedestrians cross *left to right* on any particular dump.
3. The pdf for the time between dumps.
4. The expected time that a randomly arriving pedestrian must wait until crossing.
5. The expected time that a randomly arriving *observer*, who is not a pedestrian, will wait until the next dump.

The last two parts to this question will provide a good motivation for our more general study of queues in Chapter 4.

We answer each part in sequence, analyzing each of the three decision rules as we proceed.

1. Let $\bar{N}_{LA}$, $\bar{N}_{LB}$, and $\bar{N}_{LC}$ be the expected number of pedestrians crossing left to right on any dump for each of the three decision rules A, B, and C, respectively.

For interdump times fixed at T minutes, we simply require the expected number of Poisson arrivals in an interval $[0, T]$, which gives

$$\boxed{\bar{N}_{LA} = \lambda_L T} \tag{2.67}$$

If the decision rule is to dump whenever there are exactly N_0 waiting pedestrians, obviously the total number of pedestrians crossing will be equal to N_0. However, the probability that any particular one is a left-to-right crossing pedestrian is $\lambda_L/(\lambda_L + \lambda_R)$, and the type of each successive pedestrian is chosen independently. Thus, we may think of each dump as N_0 independent Bernoulli trials each having "success" probability $\lambda_L/(\lambda_L + \lambda_R)$. Clearly, the expected number of left-to-right crossing pedestrians in this case is

$$\boxed{\bar{N}_{LB} = \frac{N_0 \lambda_L}{\lambda_L + \lambda_R}} \tag{2.68}$$

For rule C, the expected number of left-to-right crossing pedestrians during the pedestrian-initiated waiting period of duration T_0 minutes is $\lambda_L T_0$. However, we must also count the first arriving pedestrian who initiates this waiting period; he has a probability equal to $\lambda_L/(\lambda_L + \lambda_R)$ of being a "left-to-right" pedestrian. Thus, using a "left-pedestrian" indicator random variable, added to a Poisson random variable, we obtain

$$\boxed{\bar{N}_{LC} = \lambda_L T_0 + \frac{\lambda_L}{\lambda_L + \lambda_R}} \tag{2.69}$$

2. Let β_A, β_B, and β_C be the three desired probabilities for decision rules A, B, and C, respectively.

For decision rule A, β_A is simply the probability of zero Poisson arrivals in a fixed interval of length T,

$$\boxed{\beta_A = e^{-\lambda_L T}} \tag{2.70}$$

For decision rule B, the probability that zero of the N_0 crossing pedestrians are left-to-right pedestrians is equal to the probability that N_0 successive independent Bernoulli trials yield a right-to-left pedestrian. Thus,

$$\boxed{\beta_B = \left(\frac{\lambda_R}{\lambda_L + \lambda_R}\right)^{N_0}} \tag{2.71}$$

For decision rule C, β_C is equal to the probability that the first arriving pedestrian is a right-to-left pedestrian *and* that no left-to-right pedestrians arrive in the following T_0 minutes. By independence, the probability of this intersection of events is equal to the product of the individual probabilities,

$$\boxed{\beta_C = \frac{\lambda_R}{\lambda_L + \lambda_R} e^{-\lambda_L T_0}} \tag{2.72}$$

3. Let the three desired pdf's be $f_{X_A}(\cdot)$, $f_{X_B}(\cdot)$, and $f_{X_C}(\cdot)$, respectively. Since we are finding the pdf's for the time between dumps and since the decision rules utilize information on the *total* number of arriving pedestrians (and rule A utilizes no information on arriving pedestrians), we need only consider in this part the pooled Poisson process having arrival rate $\lambda_L + \lambda_R$.

The first pdf is simply the unit impulse located at T; thus,

$$\boxed{f_{X_A}(x) = \mu_0(x - T)} \tag{2.73}$$

The second pdf corresponds to the N_0th-order interarrival time pdf for a Poisson process with (pooled) arrival rate $\lambda_L + \lambda_R$. By (2.57), this is simply the N_0th-order Erlang pdf,

$$\boxed{f_{X_B}(x) = \frac{(\lambda_L + \lambda_R)^{N_0} x^{N_0 - 1} e^{-(\lambda_L + \lambda_R)x}}{(N_0 - 1)!} \qquad x \geq 0} \tag{2.74}$$

For the third pdf we use the fact that the time between dumps X_C can be expressed as

$$X_C = \text{time until first pedestrian arrives} + T_0$$

But the time until the first pedestrian arrives is equal to the first-order interarrival time in a (pooled) process with rate $\lambda_L + \lambda_R$, and this is simply a negative exponential pdf with mean $(\lambda_L + \lambda_R)^{-1}$. Thus, the pdf for X_C is simply the negative exponential pdf shifted to the right by T_0 minutes,

$$\boxed{f_{X_C}(x) = (\lambda_L + \lambda_R) e^{-(\lambda_L + \lambda_R)(x - T_0)} \qquad x \geq T_0} \tag{2.75}$$

4. Let $\bar{W}_A$, $\bar{W}_B$, and $\bar{W}_C$ be the three desired expected waiting times of a randomly arriving pedestrian. The reasoning here, while the most advanced in the problem, will provide some beginning insights into the theory of

queues, which we study in earnest in Chapters 4 and 5. As in part 3, here we need only study the pooled process having rate $\lambda_L + \lambda_R$.

For rule A, in which a dump occurs every T minutes, we can correctly reason that his or her arrival time in the interval $[0, T]$ is uniformly distributed over the interval, and thus the waiting time until the end of the interval is uniformly distributed, implying

$$\boxed{\bar{W}_A = \frac{T}{2}} \tag{2.76}$$

To analyze rule B we must be a bit more careful. A randomly arriving pedestrian is equally likely to be the first to arrive since the last dump, the second to arrive, . . . , the N_0th to arrive. Thus, the probability that a randomly arriving pedestrian is the kth to arrive since the last dump is equal to $1/N_0$ for $k = 1, 2, \ldots, N_0$. Given that he or she is the kth to arrive, his or her conditional waiting time until the next dump is distributed as an $(N_0 - k)$-order Erlang random variable with parameter $\lambda_L + \lambda_R$ (where we define a zero-order Erlang random variable to always assume the experimental value zero). Thus, the conditional mean waiting time of the kth arriving pedestrian is

$$\frac{N_0 - k}{\lambda_L + \lambda_R}$$

If we uncondition by multiplying by the probability of being the kth arriving pedestrian and summing over all possibilities, we have

$$\bar{W}_B = \sum_{k=1}^{N_0} \left(\frac{N_0 - k}{\lambda_L + \lambda_R}\right) \frac{1}{N_0} \tag{2.77}$$

If we utilize the fact that

$$\sum_{k=1}^{N_0} k = \frac{N_0(N_0 + 1)}{2} \tag{2.78}$$

(2.77) simplifies to

$$\boxed{\bar{W}_B = \frac{N_0 - 1}{2(\lambda_L + \lambda_R)}} \tag{2.79}$$

Does this make sense intuitively?

To analyze rule C, let us condition on the total number of pedestrians who arrive during the interval of length T_0 initiated by the first arriving pedestrian. Clearly, the first arriving pedestrian has an expected wait of T_0 minutes and any other arriving pedestrian prior to the dump has an expected wait of

$T_0/2$ minutes. We seek to combine these two conditional expected waits in an appropriate manner. Call $(\bar{W}_C|k)$ the conditional expected wait of a randomly arriving pedestrian, *given* that exactly k pedestrians arrive in the associated interval of length T_0. Then clearly

$$(\bar{W}_C|k) = \frac{1}{k+1}T_0 + \frac{k}{k+1}\frac{T_0}{2} = \frac{2+k}{2(k+1)}T_0$$

To find now the expected unconditional waiting time, $\bar{W}_c$, per pedestrian, we must multiply by the probability that a randomly chosen pedestrian has crossed the street in a group of $k+1$ pedestrians. This is a case of random incidence and the probability in question [use the discrete analogue of (2.63)] is given by

$$\frac{(k+1)[(\lambda_L+\lambda_R)T_0]^k e^{-(\lambda_L+\lambda_R)T_0}}{k![1+(\lambda_L+\lambda_R)T_0]} \qquad k = 0, 1, 2, \ldots$$

We then have for the unconditional expectation:

$$\bar{W}_C = \sum_{k=0}^{\infty} \frac{(2+k)T_0}{2(k+1)} \frac{(k+1)[(\lambda_L+\lambda_R)T_0]^k e^{-(\lambda_L+\lambda_R)T_0}}{k![1+(\lambda_L+\lambda_R)T_0]}$$

which simplifies to

$$\boxed{\bar{W}_C = \frac{T_0}{2}\left[1 + \frac{1}{1+(\lambda_L+\lambda_R)T_0}\right]} \tag{2.80}$$

We can check the reasonableness of this result for two limiting cases: (1) as T_0 becomes large compared to the mean passenger interarrival time $(\lambda_L+\lambda_R)^{-1}$, the mean wait approaches one half the mean interdump time, a result in agreement with (2.66) for the case in which the variance of the interdump time is small compared to the square of the mean; and (2) as T_0 becomes very small, $\bar{W}_C$ approaches T_0 (which is also expected, since in that case nearly all crossing pedestrians are first arriving pedestrians).

We could obtain (2.80) by another argument, based on "perturbation random variables," as formalized in Section 3.8. Roughly, the argument goes as follows. Each pedestrian must incur an average wait of $T_0/2$ minutes. In addition, there is a certain probability that a randomly arriving pedestrian must incur an additional mean wait of $T_0/2$ minutes; this probability is equal to the probability of being the first arriving pedestrian. Thus, we could write $\bar{W}_C$ as the sum of $T_0/2$ plus a perturbation term, as follows:

$$\bar{W}_C = \frac{T_0}{2} + \frac{T_0}{2}P\{\text{random pedestrian first to arrive since last dump}\} \tag{2.81}$$

To compute the probability in the perturbation term, consider a very long period of time during which N dumps, where N is a large number, have taken place. This means that N "first pedestrians" have arrived and an expected number of $N(\lambda_L + \lambda_R)T_0$ pedestrians arrived during the T_0-long intervals that follow the arrival of the first pedestrian. Thus,

$$\begin{aligned} P\{\text{randomly chosen pedestrian was first to arrive since last dump}\} &\approx \frac{N}{N + N(\lambda_L + \lambda_R)T_0} \\ &= \frac{1}{1 + (\lambda_L + \lambda_R)T_0} \end{aligned}$$

Substituting into (2.81), we again obtain

$$\bar{W}_C = \frac{T_0}{2}\left[1 + \frac{1}{1 + (\lambda_L + \lambda_R)T_0}\right]$$

Note that the result above has been argued quite informally, in keeping with the informality of the perturbation argument.

5. This part is considerably easier than part 4. However, it is instructive to compare the answers to those in part 4 for each of the three decision rules. In part 4, the randomly arriving person is a pedestrian and he or she thus "disturbs" the system. In this part the randomly arriving person is strictly an observer, and thus he or she does not affect the system. Call the times desired $\bar{V}_A$, $\bar{V}_B$, and $\bar{V}_C$ for the three respective decision rules.

All we need do to obtain the three required answers is apply (2.66) derived for random incidence. Recalling that equation, in the context of this problem, it states that the expected time until the next dump (from the moment of the observer's arrival) is equal to one half the sum of the mean squared and the variance of the interdump times, divided by the mean. For instance, for decision rule A we have $E[X_A] = T$ and $\sigma^2_{X_A} = 0$, implying that

$$\boxed{\bar{V}_A = \frac{T}{2}} \tag{2.82}$$

as expected.

For decision rule B, we have $E[X_B] = N_0/(\lambda_L + \lambda_R)$ and $\sigma^2_{X_B} = N_0/(\lambda_L + \lambda_R)^2$, implying that

$$\boxed{\bar{V}_B = \frac{N_0 + 1}{2(\lambda_L + \lambda_R)}} \tag{2.83}$$

Note that this result is similar to that derived in part 4, except that the "+1" here is replaced by a "−1." Does this make sense intuitively? Can you derive (2.83) along the lines argued in part 4 to confirm this result?

For decision rule C, we have

$$E[X_C] = \frac{1}{\lambda_L + \lambda_R} + T_0 \qquad \text{and} \qquad \sigma^2_{X_C} = \left(\frac{1}{\lambda_L + \lambda_R}\right)^2$$

Thus,

$$\boxed{\bar{V}_C = \frac{T_0}{2} + \frac{1}{2(\lambda_L + \lambda_R)} + \frac{1}{2(\lambda_L + \lambda_R)(1 + T_0[\lambda_L + \lambda_R])}} \tag{2.84}$$

This result, too, checks with our intuition in two limiting cases: as $T_0 \longrightarrow 0$, $\bar{V}_C \longrightarrow 1/(\lambda_L + \lambda_R)$, implying that the expected time until the next dump is equal to the expected time of arrival of the next pedestrian. As T_0 becomes large, $\bar{V}_C \longrightarrow [T_0 + (\lambda_L + \lambda_R)^{-1}]/2$, which is one half the expected interdump time; this result checks with (2.66) for the case in which the variance of the interdump time is negligibly small in comparison to the square of the mean.

This completes our analysis of the pedestrian crossing problem. In addition to providing us with a good example of probabilistic modeling as applied to an urban service system, it has pointed to directions we wish to pursue in Chapters 3–5.

Example 8: Pedestrian Crossing, One More Time

One way to evaluate the alternative decision rules is to design each system to yield the same mean pedestrian waiting time and then to determine which system has the smallest average frequency of pedestrian light flashes (a small frequency would reduce the disruption of vehicular traffic). Or, we could hold constant the frequency of pedestrian light flashes and determine which system has the smallest mean pedestrian waiting time. Intuitively, how would you rank-order the systems according to this evaluation yardstick? Can you prove your intuition mathematically? Try several typical values of λ_L, λ_R and desired mean wait to observe the range and sensitivity of the results.

Hint: Write an expression for the mean time between pedestrian light flashes for each of the three systems. You should then be able to express each of the pedestrian mean waiting times derived in part (4) in the form $\frac{1}{2}\cdot$(mean time between pedestrian light flashes)—(perturbation term). By examining the perturbation terms, you should be able to show that system A is the least preferred, system B the most preferred, and system C falls between the two. However, in comparing systems B and C, you should also consider the technological feasibility and associated cost of each of the systems.

2.15 CONCLUSION

This chapter has reviewed the fundamentals of probabilistic modeling, with emphasis on modeling physical situations in an urban setting. In practice there is no magic procedure for developing such models. In fact, it is the reduction of a situation to a sample space with its probability assignment and random variables that is usually the most difficult step. The formal manipulation that follows from that point is (relatively) straightforward but still quite interesting. By having described physical situations in terms of word statements, we hope to have given the reader some feeling for the problem definition process. However, our problems in this book are preselected to illustrate concepts relevant to that portion of the text. They are also usually selected to yield a tractable analysis. No such guarantee is available in an actual city, and one is hard-pressed to find problems tightly worded with no ambiguity. But these complexities contribute to the challenge of urban analysis, which in many ways is as much an art form as a formalism.

Armed with the prerequisite background of Chapters 1 and 2, we now proceed to methods, procedures, and points of view in modeling analysis that are especially important in an urban setting.

Problems

The exercises contained within this chapter have been designed to give the reader needed review experience in formal manipulation in probabilistic analysis. The following four problems have the dual purpose of continuing such review and providing experience in the modeling of physical situations that typically arise in urban analysis. The reader is also referred to Problem 3.1 for more review on the basic concepts of probability modeling.

2.1 *Discrete random variables: a clean sweep* A 300-foot-long city block face contains eight 25-foot-long parking spaces, as shown in Figure P2.1. For convenience, we have numbered the parking spaces consecutively from 1 to 8. Each Tuesday at 8:00 A.M. a street sweeper attempts to clean the entire street. Any cars parked in those spaces inhibit the work of the sweeper and are, accordingly, each given a $25 fine (parking ticket). We assume that the probability that any given parking space will be occupied by an illegally parked car is p and that the status (full or empty) of each parking space is independent of the status of all other parking spaces.

a. For any given Tuesday morning, determine the mean and variance of the number of illegally parked cars on this block face.

b. Repeat part (a) for the mean and variance of the total dollar value of parking tickets issued.

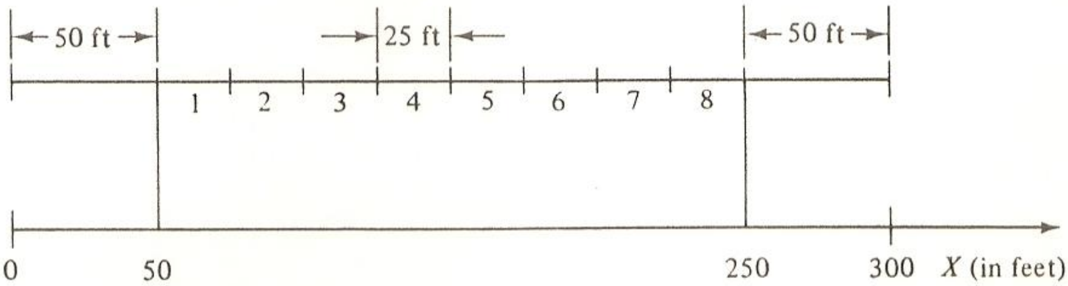
For each parked car, the sweeper is unable to sweep the 25 feet contained within that car's parking space and, in addition, because of maneuverability problems, it is unable to sweep within 12.5 feet on either side of that car's parking space (we ignore size differences of cars here). For instance, if cars were only parked in spaces 1 and 3, the sweeper would be unable to sweep 100 consecutive feet of the street, starting at $x = 37.5$ feet (see Figure P2.1).

c. Determine the probability of a clean sweep of the street (i.e., that no illegally parked cars are present).

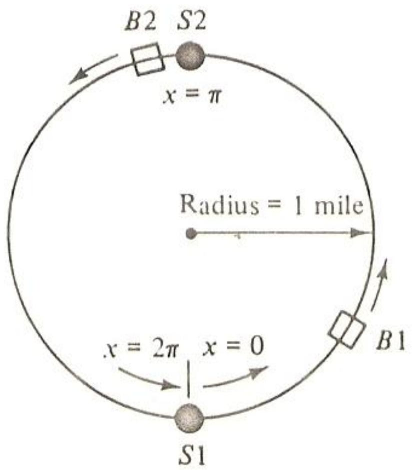
d. Given that there are exactly two illegally parked cars on the street, determine the conditional pmf for the length (in feet) of street swept.

e. Find the unconditional mean length of street swept.

Hint: Avoid "brute-force" methods by arguing as follows: for any "interior" point x, $62.5 < x < 237.5$, the probability that x will be swept is $(1 - p)^2$. Why? Then, the expected length of parking space 4 that will be swept is $25(1 - p)^2$. Continuing this reasoning and including the other "noninterior" points, you should find that the answer is $75 + 50(1 - p) + 175(1 - p)^2$ feet. (This line of reasoning utilizes ideas of coverage that are explained more fully in Chapter 3.)



2.2 *Mixed random variables: a shuttle* Suppose that two buses, $B1$ and $B2$, are used to shuttle passengers between two stations, $S1$ and $S2$, as shown in Figure P2.2. For convenience, we denote the location x_1 and x_2 of each bus $B1$ and $B2$, respectively, by its counterclockwise distance from $S1$, $0 \leq x < 2\pi$. We are given the following facts about system operation:



1. Statistically, buses $B1$ and $B2$ behave identically, as do stations $S1$ and $S2$.
2. Each bus spends 25 percent of its time at station $S1$, 25 percent at $S2$, and 50 percent traveling between the two stations.
3. Speed of travel is a fixed constant.

a. Draw the cdf and pdf for the location of $B1$ at a randomly chosen time. [If the cdf is differentiated to yield the pdf, any impulses that result (associated with probability masses) are usually sketched as spikes with numbers adjacent to the spikes equal to the corresponding probability masses.]

For parts (b) and (c), assume that $B1$ and $B2$ operate completely independently. (Is that reasonable?)

b. Sketch and describe the joint pdf for the locations of $B1$ and $B2$. (You may have to extend to two dimensions your notion of impulses.)

c. Define the *headway* H between $B1$ and $B2$ as the clockwise distance from $B1$ to $B2$.

   i. Argue that

$$H = \begin{cases} x_1 - x_2 & \text{if } x_1 \geq x_2 \\ x_1 - x_2 + 2\pi & \text{if } x_1 < x_2 \end{cases}$$

   For the remaining parts in (c), let

$$A = \text{event that } \left\{H \leq \frac{\pi}{2}\right\}$$

   ii. Sketch the conditional joint pdf for x_1 and x_2, given A.

   iii. Sketch the conditional marginal pdf for x_1, given A.

   iv. Find the conditional mean headway, given A

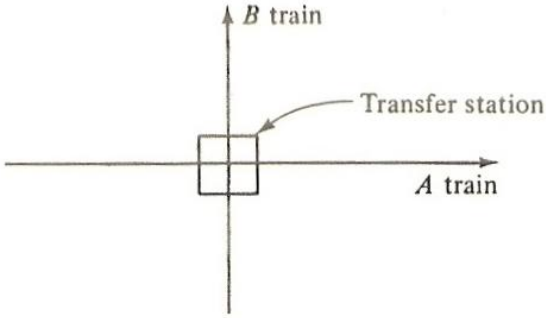
$$\left(\text{i.e., } E[H \mid A] = E\left[H \mid H \leq \frac{\pi}{2}\right]\right)$$

d. Suppose that the system operates mechanically so that

$$x_2 = \begin{cases} x_1 + \pi & \text{whenever } 0 \leq x_1 < \pi \\ x_1 - \pi & \text{whenever } \pi \leq x_1 < 2\pi \end{cases}$$

   Sketch and describe the joint pdf of x_1 and x_2.

2.3 *Poisson processes: subways are for waiting.* Two one-way subway lines, the A train line and the B train line, intersect at a transfer station, as shown in Figure P2.3. A trains and B trains arrive at the station according to independently operating



Poisson processes, with Poisson rates of arrival

$$\lambda_A = 3\ A \text{ trains per hour}$$
$$\lambda_B = 6\ B \text{ trains per hour}$$

We assume that passenger boarding and unboarding occurs almost instantaneously, not unlike true rush-hour conditions in many cities throughout the world!

a. At a random time, Bart, a prospective passenger, arrives at the station, awaiting an A train.

   i. What is the pdf for the time he will have to wait?

   ii. What is the probability that *at least* 3 B trains will arrive while Bart is waiting?

   iii. What is the probability that exactly 3 B trains will arrive while Bart is waiting?

b. What is the probability that the station handles exactly 9 trains during any given hour?

c. If an observer counts the number of trains that the station handles each hour, starting at 8:00 A.M. on Tuesday, what is the expected number of hours until he or she will *first* count exactly 9 trains during an hour that commences "on the hour"? (e.g., 9:00 A.M., 10:00 A.M., 2:00 P.M.)

d. Whenever an A train is ready to depart from the station, it will be held if an approaching B train is within 30 seconds of the station. This delay policy is to facilitate the rapid transfer of passengers from the B train to the A train.

   i. Approximately what fraction of A trains are delayed in this manner?

   ii. Given a B-train passenger who benefits from such instantaneous transfer to an A train, using the up-to-30-second-delay policy for the A train, compute his or her mean waiting time reduction and compare to the mean increase in travel time for an already boarded A train passenger. Intuitively, under what conditions would such a delay policy provide net global travel-time reduction?