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 .... 1, 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 ⊘, for universal event. An event containing no points in the sample space is called o, 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, . . . , 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,...,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 I 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 A 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 1/2:

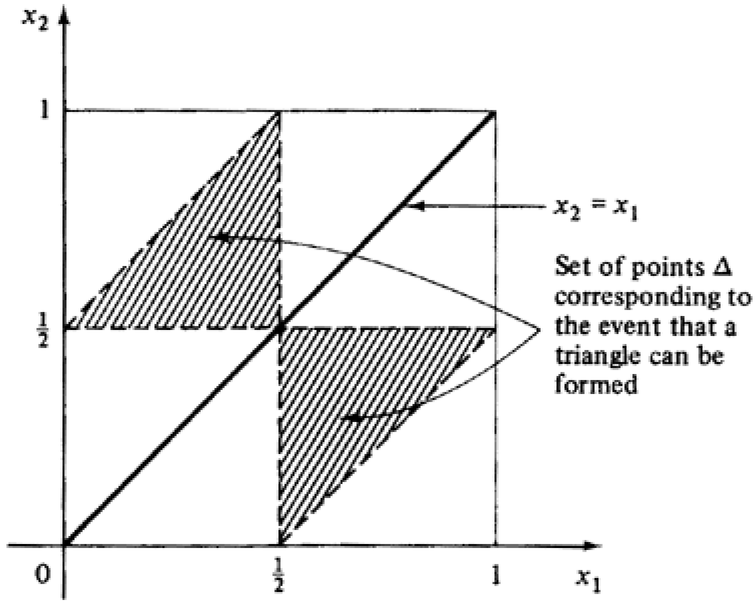

FIGURE 2.1 Sample space for the broken-stick experiment.

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

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

The resulting set is contained in the lower triangle of area 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 1/8 above the line $x_1 = x_2$.

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

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

then

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

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).

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 inA (i.e.,A⊓C = C or A ⊓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 eventsA_1 and A_2, $P\{B|A_1 \sqcap A_2\}$ 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

$$P\{A \cap B\} = P\{A\}P\{B\} \qquad (2.2)$$

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, 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, 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 Aj, j ≠ 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 triangleproblem, 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 thatX = 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 px(·) is the probability mass function (pmf) of the (discrete) random varialble 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,

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

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 sincePx(∞) = 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:

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

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,

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

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,

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

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 passenger (n = 1, 2,..., 5), each possible combination of male and female passenger 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 mini bus.
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:

$N_F \equiv$ **number of females on the minibus**
$N_M \equiv$ **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 = event that i total passengers are on the bus

We know that $P\{A_1\} = 1/5$, i = 1, 2,. .., 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

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

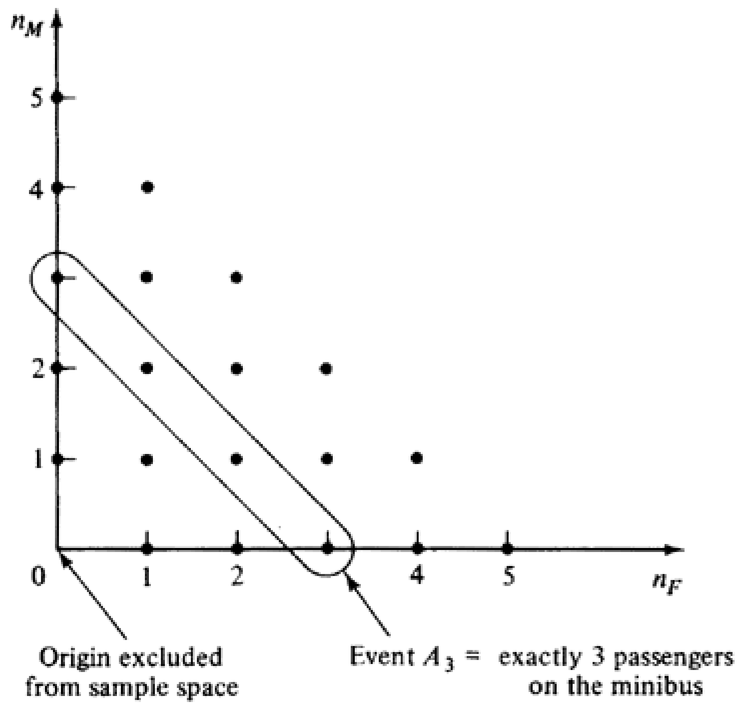

FIGURE 2.2 Sample space for minibus problem.

This is the answer to part (a). It says, roughly, that the probability that $\{N_f = n_f$ and $N_m = n_M\}$ is equal to 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 $P\{N_f = 3) = P_{Nf}(3)$, we sum the probabilities corresponding to the (finest-grained) events $\{N_f = 3, N_m = 0)$, $\{N_f = 3, Nm = 1\}$, and $\{N_f = 3, N_m = 2\}$, yielding 1/20 + 1/25 + 1/30 = 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 As (i.e., $n_F + n_m = 5$). Thus, invoking (2.5),

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

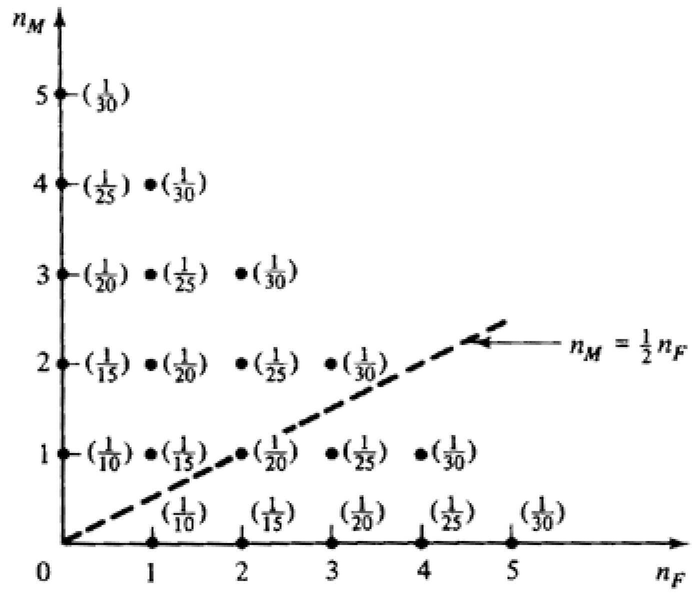

FIGURE 2.3 Joint pmf for minibus problem.

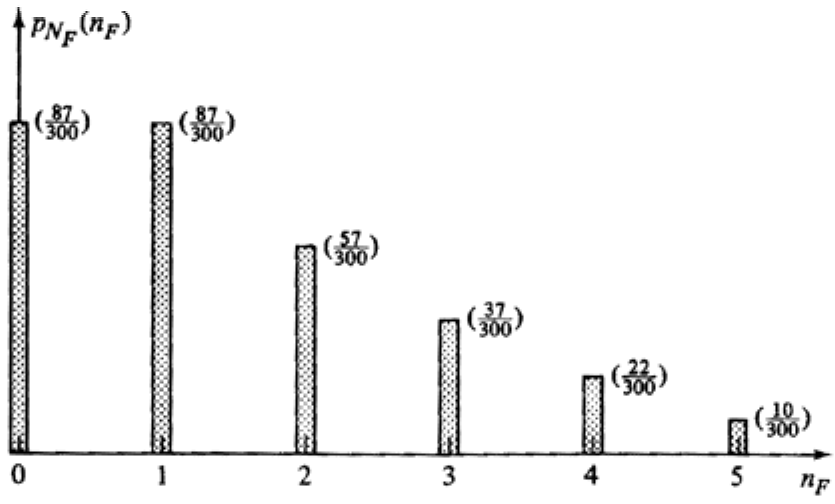

FIGURE 2.4 PMF for the number of female minibus passengers.

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 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 $P_{Nf}(n_f \mid B)$. 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 $n_m = I\ n_F$ (shown in Figure 2.3). Summing the probabilities of the eight finestgrained outcomes satisfying this inequality, we find that P{B} = 124/300 = 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 1/31/75. 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).

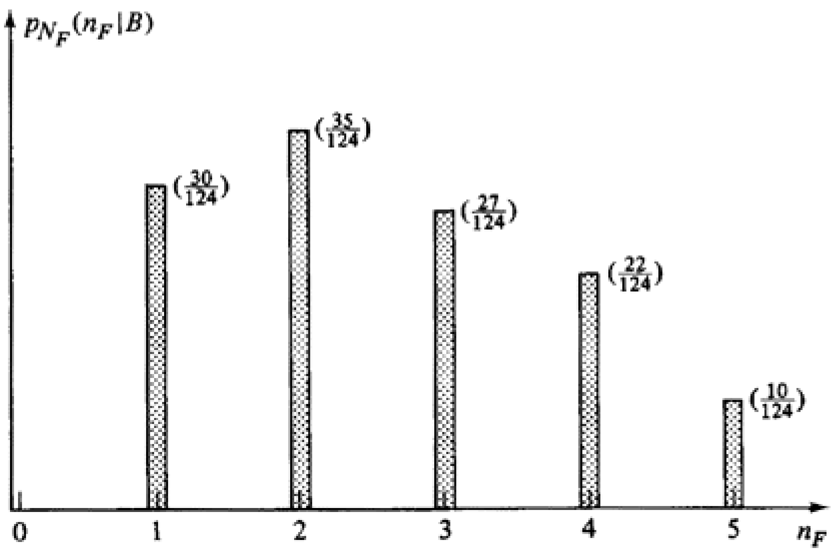

FIGURE 2.5 Conditional marginal pmf for the number of female passengers, given *B*.

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

$p_{x|y}(x|y) = p_x(x)$

and

$p_{x,y}(x,y) = p_x(x)p_y(y)$ for all possible values of x and y (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,

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

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?

2.7 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

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

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

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

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(·) 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 = I results in an expected value E[X] = 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,

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

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

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

[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.

2.9.1 Bernoulli PMF

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

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

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

The mean and variance of a Bernoulli random variable are

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

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

The z-transform is $p^t_x(z) = (l - p) + pz$. The Bernoulli pmf arises in simple trials having only two outcomes; it is also useful in the analysis of setindicator 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 I 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

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

The z-transform is

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

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 ....) 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 (I - 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 σ_X = 2 $\sqrt{3}$ ≈ 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

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

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

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

Recalling the binomial theorem,

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

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:

$$E[W] = np \qquad (2.30a)$$

$$\sigma_W^2 = np(1-p) \qquad (2.30b)$$

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

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

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

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

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

$$E[K] = \mu \qquad (2.33a)$$

$$\sigma_K^2 = \mu \qquad (2.33b)$$

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, . . . , Poisson-type events occur (see Figure 2.6). Suppose that

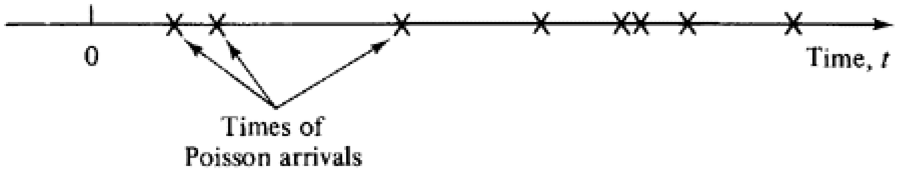

FIGURE 2.6 Poisson arrivals in time.

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:

fx(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, . . . , 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),

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

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

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

Example 3: Uniform PDF

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

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

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.

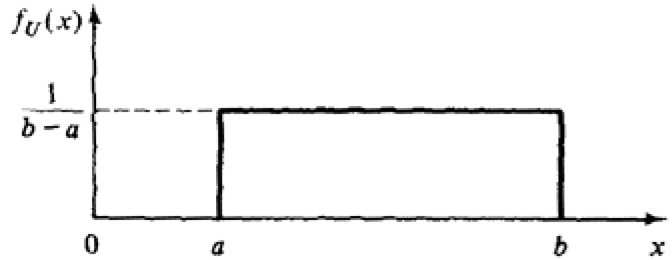

FIGURE 2.7 PDF of a uniformly distributed random variable.

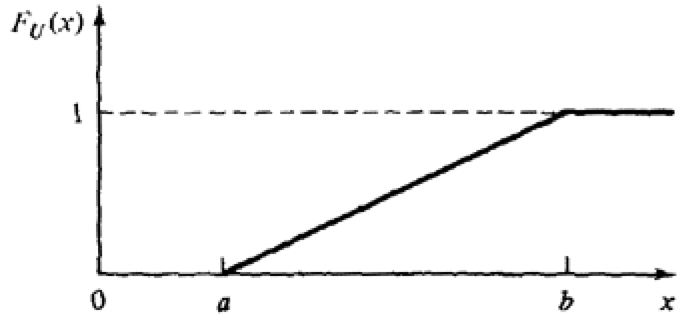

FIGURE 2.8 CDF of a uniformly distributed random variable.

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,

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

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 $X_i, f_{xi}(x_i)$, defined so that $f_{xi}(x_i)dx_i \equiv P\{x_i \leq X_i \leq x_i + dx_i\}$. We can calculate the marginal from the joint pdf simply by integrating over all the other random variables.

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

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

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

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

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

or, similarly,

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

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

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

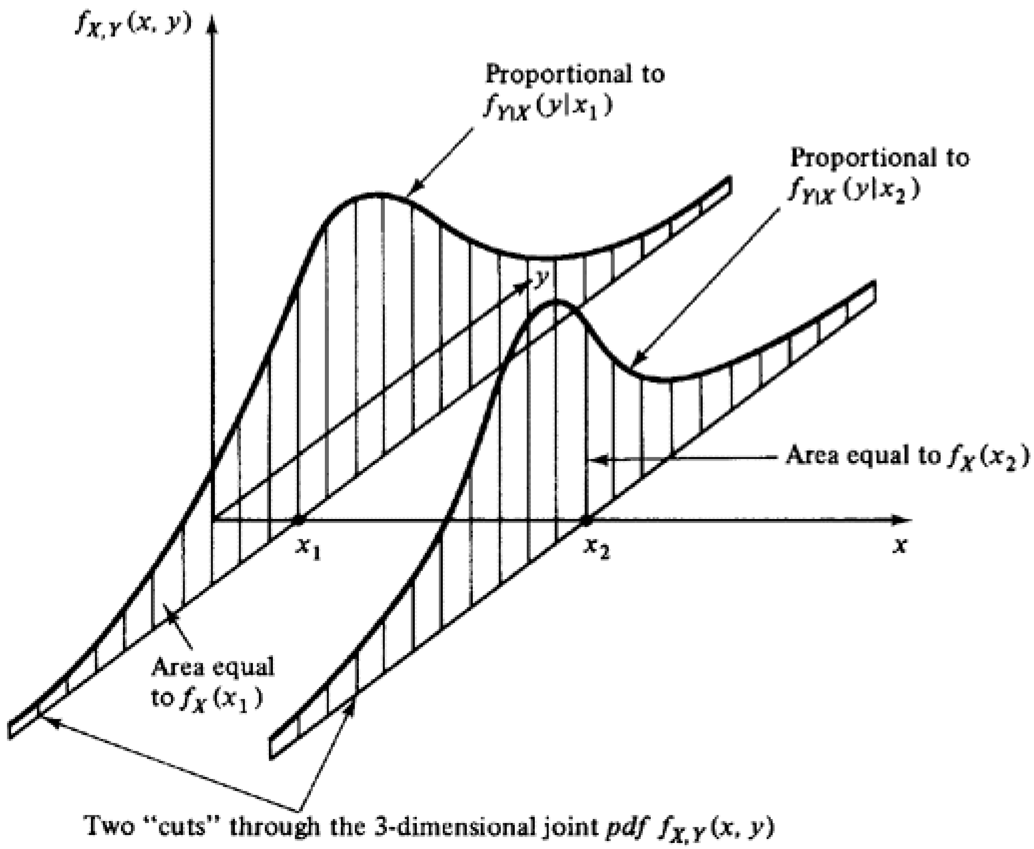

FIGURE 2.9 Pictorial depiction of $f_{Y|X}(y \mid x)$.

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 = 1/4 and a tri-angle 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

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

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 1/8 and since the joint pdf is uniform with height 1, we obtain by inspection that P{Δ} = 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 = 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 $f_{X_1,X_2}(x_1, x_2 \mid \Delta)$ = 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 = 1/4 and a triangle cannot be formed. By inspection of Figure 2.1, the conditional pdf for X_2 is

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

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

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

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

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

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

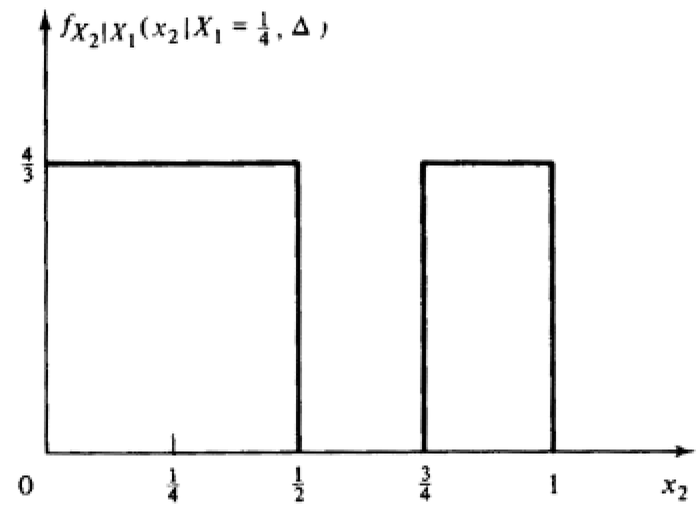

FIGURE 2.10 Conditional PDF for triangle problem.

or the conditional variance

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

2.11.1 Uniform PDF

In Section 2. 10 we introduced the uniform pdf,

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

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

$$E[U] = \frac{a+b}{2} \qquad (2.48a)$$

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

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

2.11.4 Gaussian PDF

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

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

The mean, variance, and s-transforms are

$$E[Y] = m_y \qquad (2.53a)$$

$$\sigma_Y^2 = \sigma^2 \qquad (2.53b)$$

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

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.

2.11.3 Erlang PDF

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

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

The mean, variance, and s-transforms are

$$E[L_k] = \frac{k}{\lambda} \qquad (2.51a)$$

$$\sigma^2_{L_k} = \frac{k}{\lambda^2} \qquad (2.51b)$$

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

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

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

The mean, variance, and s-transforms are

$$E[Y] = m_y \qquad (2.53a)$$

$$\sigma^2_Y = \sigma^2 \qquad (2.53b)$$

$$f^T_Y(s) = e^{-m_y s + (1/2)\sigma^2 s^2} \qquad (2.53c)$$

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.