Introduction to Probability Theory

1.1 Introduction

Any realistic model of a real-world phenomenon must take into account the possibility of randomness. That is, more often than not, the quantities we are interested in will not be predictable in advance but, rather, will exhibit an inherent variation that should be taken into account by the model. This is usually accomplished by allowing the model to be probabilistic in nature. Such a model is, naturally enough, referred to as a probability model.

The majority of the chapters of this book will be concerned with different probability models of natural phenomena. Clearly, in order to master both the "model building" and the subsequent analysis of these models, we must have a certain knowledge of basic probability theory. The remainder of this chapter, as well as the next two chapters, will be concerned with a study of this subject.

1.2 Sample Space and Events

Suppose that we are about to perform an experiment whose outcome is not predictable in advance. However, while the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known. This set of all possible outcomes of an experiment is known as the *sample space* of the experiment and is denoted by S.

Introduction to Probability Models, ISBN: 9780123756862
Copyright © 2010 by Elsevier, Inc. All rights reserved.

Some examples are the following.

1. If the experiment consists of the flipping of a coin, then

$$S = \{H, T\}$$

where H means that the outcome of the toss is a head and T that it is a tail.

2. If the experiment consists of rolling a die, then the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}$$

where the outcome i means that i appeared on the die, $i = 1, 2, 3, 4, 5, 6$.

3. If the experiments consists of flipping two coins, then the sample space consists of the following four points:

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

The outcome will be (H, H) if both coins come up heads; it will be (H, T) if the first coin comes up heads and the second comes up tails; it will be (T, H) if the first comes up tails and the second heads; and it will be (T, T) if both coins come up tails.

4. If the experiment consists of rolling two dice, then the sample space consists of the following 36 points:

$$S = \begin{Bmatrix} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{Bmatrix}$$

where the outcome (i, j) is said to occur if i appears on the first die and j on the second die.

5. If the experiment consists of measuring the lifetime of a car, then the sample space consists of all nonnegative real numbers. That is,

$$S = [0, \infty)^* \qquad \blacksquare$$

Any subset E of the sample space S is known as an *event*. Some examples of events are the following.

1′. In Example (1), if $E = \{H\}$, then E is the event that a head appears on the flip of the coin. Similarly, if $E = \{T\}$, then E would be the event that a tail appears.

2′. In Example (2), if $E = \{1\}$, then E is the event that one appears on the roll of the die. If $E = \{2, 4, 6\}$, then E would be the event that an even number appears on the roll.

* The set (a, b) is defined to consist of all points x such that $a < x < b$. The set $[a, b]$ is defined to consist of all points x such that $a \leqslant x \leqslant b$. The sets $(a, b]$ and $[a, b)$ are defined, respectively, to consist of all points x such that $a < x \leqslant b$ and all points x such that $a \leqslant x < b$.

3′. In Example (3), if $E = \{(H, H),\ (H, T)\}$, then E is the event that a head appears on the first coin.
4′. In Example (4), if $E = \{(1, 6),\ (2, 5),\ (3, 4),\ (4, 3),\ (5, 2),\ (6, 1)\}$, then E is the event that the sum of the dice equals seven.
5′. In Example (5), if $E = (2, 6)$, then E is the event that the car lasts between two and six years. ■

We say that the event E occurs when the outcome of the experiment lies in E. For any two events E and F of a sample space S we define the new event $E \cup F$ to consist of all outcomes that are either in E or in F or in both E and F. That is, the event $E \cup F$ will occur if *either* E or F occurs. For example, in (1) if $E = \{H\}$ and $F = \{T\}$, then

$$E \cup F = \{H,\ T\}$$

That is, $E \cup F$ would be the whole sample space S. In (2) if $E = \{1, 3, 5\}$ and $F = \{1, 2, 3\}$, then

$$E \cup F = \{1, 2, 3, 5\}$$

and thus $E \cup F$ would occur if the outcome of the die is 1 or 2 or 3 or 5. The event $E \cup F$ is often referred to as the *union* of the event E and the event F.

For any two events E and F, we may also define the new event EF, sometimes written $E \cap F$, and referred to as the *intersection* of E and F, as follows. EF consists of all outcomes which are *both* in E and in F. That is, the event EF will occur only if both E and F occur. For example, in (2) if $E = \{1, 3, 5\}$ and $F = \{1, 2, 3\}$, then

$$EF = \{1, 3\}$$

and thus EF would occur if the outcome of the die is either 1 or 3. In Example (1) if $E = \{H\}$ and $F = \{T\}$, then the event EF would not consist of any outcomes and hence could not occur. To give such an event a name, we shall refer to it as the null event and denote it by Ø. (That is, Ø refers to the event consisting of no outcomes.) If $EF = Ø$, then E and F are said to be *mutually exclusive*.

We also define unions and intersections of more than two events in a similar manner. If $E_1, E_2, \ldots$ are events, then the union of these events, denoted by $\bigcup_{n=1}^{\infty} E_n$, is defined to be the event that consists of all outcomes that are in E_n for at least one value of $n = 1, 2, \ldots$. Similarly, the intersection of the events E_n, denoted by $\bigcap_{n=1}^{\infty} E_n$, is defined to be the event consisting of those outcomes that are in all of the events $E_n, n = 1, 2, \ldots$.

Finally, for any event E we define the new event E^c, referred to as the *complement* of E, to consist of all outcomes in the sample space S that are not in E. That is, E^c will occur if and only if E does not occur. In Example (4) if $E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$, then E^c will occur if the sum of the dice does not equal seven. Also note that since the experiment must result in some outcome, it follows that $S^c = \varnothing$.

1.3 Probabilities Defined on Events

Consider an experiment whose sample space is S. For each event E of the sample space S, we assume that a number $P(E)$ is defined and satisfies the following three conditions:

(i) $0 \leqslant P(E) \leqslant 1$.
(ii) $P(S) = 1$.
(iii) For any sequence of events $E_1, E_2, \dots$ that are mutually exclusive, that is, events for which $E_n E_m = \varnothing$ when $n \neq m$, then

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

We refer to $P(E)$ as the probability of the event E.

Example 1.1 In the coin tossing example, if we assume that a head is equally likely to appear as a tail, then we would have

$$P(\{H\}) = P(\{T\}) = \tfrac{1}{2}$$

On the other hand, if we had a biased coin and felt that a head was twice as likely to appear as a tail, then we would have

$$P(\{H\}) = \tfrac{2}{3}, \qquad P(\{T\}) = \tfrac{1}{3}$$ ■

Example 1.2 In the die tossing example, if we supposed that all six numbers were equally likely to appear, then we would have

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \tfrac{1}{6}$$

From (iii) it would follow that the probability of getting an even number would equal

$$\begin{aligned} P(\{2,4,6\}) &= P(\{2\}) + P(\{4\}) + P(\{6\}) \\ &= \tfrac{1}{2} \end{aligned}$$ ■

Remark We have chosen to give a rather formal definition of probabilities as being functions defined on the events of a sample space. However, it turns out that these probabilities have a nice intuitive property. Namely, if our experiment is repeated over and over again then (with probability 1) the proportion of time that event E occurs will just be $P(E)$.

Since the events E and E^c are always mutually exclusive and since $E \cup E^c = S$ we have by (ii) and (iii) that

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

or

$$P(E^c) = 1 - P(E) \tag{1.1}$$

In words, Equation (1.1) states that the probability that an event does not occur is one minus the probability that it does occur.

We shall now derive a formula for $P(E \cup F)$, the probability of all outcomes either in E or in F. To do so, consider $P(E) + P(F)$, which is the probability of all outcomes in E plus the probability of all points in F. Since any outcome that is in both E and F will be counted twice in $P(E) + P(F)$ and only once in $P(E \cup F)$, we must have

$$P(E) + P(F) = P(E \cup F) + P(EF)$$

or equivalently

$$P(E \cup F) = P(E) + P(F) - P(EF) \tag{1.2}$$

Note that when E and F are mutually exclusive (that is, when $EF = \varnothing$), then Equation (1.2) states that

$$\begin{aligned} P(E \cup F) &= P(E) + P(F) - P(\varnothing) \\ &= P(E) + P(F) \end{aligned}$$

a result which also follows from condition (iii). (Why is $P(\varnothing) = 0$?)

Example 1.3 Suppose that we toss two coins, and suppose that we assume that each of the four outcomes in the sample space

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

is equally likely and hence has probability $\frac{1}{4}$. Let

$$E = \{(H, H), (H, T)\} \quad \text{and} \quad F = \{(H, H), (T, H)\}$$

That is, E is the event that the first coin falls heads, and F is the event that the second coin falls heads.

By Equation (1.2) we have that $P(E \cup F)$, the probability that either the first or the second coin falls heads, is given by

$$\begin{aligned}P(E \cup F) &= P(E) + P(F) - P(EF)\\ &= \tfrac{1}{2} + \tfrac{1}{2} - P(\{H, H\})\\ &= 1 - \tfrac{1}{4} = \tfrac{3}{4}\end{aligned}$$

This probability could, of course, have been computed directly since

$$P(E \cup F) = P(\{H, H), (H, T), (T, H)\}) = \tfrac{3}{4} \qquad \blacksquare$$

We may also calculate the probability that any one of the three events E or F or G occurs. This is done as follows:

$$P(E \cup F \cup G) = P((E \cup F) \cup G)$$

which by Equation (1.2) equals

$$P(E \cup F) + P(G) - P((E \cup F)G)$$

Now we leave it for you to show that the events $(E \cup F)G$ and $EG \cup FG$ are equivalent, and hence the preceding equals

$$\begin{aligned}&P(E \cup F \cup G)\\ &\quad = P(E) + P(F) - P(EF) + P(G) - P(EG \cup FG)\\ &\quad = P(E) + P(F) - P(EF) + P(G) - P(EG) - P(FG) + P(EGFG)\\ &\quad = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG) \qquad (1.3)\end{aligned}$$

In fact, it can be shown by induction that, for any n events $E_1, E_2, E_3, \dots, E_n$,

$$\begin{aligned}P(E_1 \cup E_2 \cup \cdots \cup E_n) = &\sum_i P(E_i) - \sum_{i<j} P(E_i E_j) + \sum_{i<j<k} P(E_i E_j E_k)\\ &- \sum_{i<j<k<l} P(E_i E_j E_k E_l)\\ &+ \cdots + (-1)^{n+1} P(E_1 E_2 \cdots E_n) \qquad (1.4)\end{aligned}$$

In words, Equation (1.4), known as the *inclusion-exclusion identity*, states that the probability of the union of n events equals the sum of the probabilities of these events taken one at a time minus the sum of the probabilities of these events taken two at a time plus the sum of the probabilities of these events taken three at a time, and so on.

1.4 Conditional Probabilities

Suppose that we toss two dice and that each of the 36 possible outcomes is equally likely to occur and hence has probability $\frac{1}{36}$. Suppose that we observe that the first die is a four. Then, given this information, what is the probability that the sum of the two dice equals six? To calculate this probability we reason as follows: Given that the initial die is a four, it follows that there can be at most six possible outcomes of our experiment, namely, $(4, 1)$, $(4, 2)$, $(4, 3)$, $(4, 4)$, $(4, 5)$, and $(4, 6)$. Since each of these outcomes originally had the same probability of occurring, they should still have equal probabilities. That is, given that the first die is a four, then the (conditional) probability of each of the outcomes $(4, 1)$, $(4, 2)$, $(4, 3)$, $(4, 4)$, $(4, 5)$, $(4, 6)$ is $\frac{1}{6}$ while the (conditional) probability of the other 30 points in the sample space is 0. Hence, the desired probability will be $\frac{1}{6}$.

If we let E and F denote, respectively, the event that the sum of the dice is six and the event that the first die is a four, then the probability just obtained is called the conditional probability that E occurs given that F has occurred and is denoted by

$$P(E|F)$$

A general formula for $P(E|F)$ that is valid for all events E and F is derived in the same manner as the preceding. Namely, if the event F occurs, then in order for E to occur it is necessary for the actual occurrence to be a point in both E and in F, that is, it must be in EF. Now, because we know that F has occurred, it follows that F becomes our new sample space and hence the probability that the event EF occurs will equal the probability of EF relative to the probability of F. That is,

$$P(E|F) = \frac{P(EF)}{P(F)} \tag{1.5}$$

Note that Equation (1.5) is only well defined when $P(F) > 0$ and hence $P(E|F)$ is only defined when $P(F) > 0$.

Example 1.4 Suppose cards numbered one through ten are placed in a hat, mixed up, and then one of the cards is drawn. If we are told that the number on the drawn card is at least five, then what is the conditional probability that it is ten?

Solution: Let E denote the event that the number of the drawn card is ten, and let F be the event that it is at least five. The desired probability is $P(E|F)$. Now, from Equation (1.5)

$$P(E|F) = \frac{P(EF)}{P(F)}$$

However, $EF = E$ since the number of the card will be both ten and at least five if and only if it is number ten. Hence,

$$P(E|F) = \frac{\frac{1}{10}}{\frac{6}{10}} = \frac{1}{6} \qquad \blacksquare$$

Example 1.5 A family has two children. What is the conditional probability that both are boys given that at least one of them is a boy? Assume that the sample space S is given by $S = \{(b,b),(b,g),(g,b),(g,g)\}$, and all outcomes are equally likely. ((b,g) means, for instance, that the older child is a boy and the younger child a girl.)

Solution: Letting B denote the event that both children are boys, and A the event that at least one of them is a boy, then the desired probability is given by

$$\begin{aligned} P(B|A) &= \frac{P(BA)}{P(A)} \\ &= \frac{P(\{(b,b)\})}{P(\{(b,b),(b,g),(g,b)\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \end{aligned} \qquad \blacksquare$$

Example 1.6 Bev can either take a course in computers or in chemistry. If Bev takes the computer course, then she will receive an A grade with probability $\frac{1}{2}$; if she takes the chemistry course then she will receive an A grade with probability $\frac{1}{3}$. Bev decides to base her decision on the flip of a fair coin. What is the probability that Bev will get an A in chemistry?

Solution: If we let C be the event that Bev takes chemistry and A denote the event that she receives an A in whatever course she takes, then the desired probability is $P(AC)$. This is calculated by using Equation (1.5) as follows:

$$\begin{aligned} P(AC) &= P(C)P(A|C) \\ &= \tfrac{1}{2}\tfrac{1}{3} = \tfrac{1}{6} \end{aligned} \qquad \blacksquare$$

Example 1.7 Suppose an urn contains seven black balls and five white balls. We draw two balls from the urn without replacement. Assuming that each ball in the urn is equally likely to be drawn, what is the probability that both drawn balls are black?

Solution: Let F and E denote, respectively, the events that the first and second balls drawn are black. Now, given that the first ball selected is black, there are six remaining black balls and five white balls, and so $P(E|F) = \frac{6}{11}$. As $P(F)$ is clearly $\frac{7}{12}$, our desired probability is

$$\begin{aligned} P(EF) &= P(F)P(E|F) \\ &= \tfrac{7}{12}\tfrac{6}{11} = \tfrac{42}{132} \end{aligned} \qquad \blacksquare$$

Hence, the probability that none of the men selects his own hat is $1-\frac{2}{3}=\frac{1}{3}$. ■

1.5 Independent Events

Two events E and F are said to be *independent* if

$$P(EF) = P(E)P(F)$$

By Equation (1.5) this implies that E and F are independent if

$$P(E|F) = P(E)$$

(which also implies that $P(F|E) = P(F)$). That is, E and F are independent if knowledge that F has occurred does not affect the probability that E occurs. That is, the occurrence of E is independent of whether or not F occurs.

Two events E and F that are not independent are said to be *dependent*.

Example 1.9 Suppose we toss two fair dice. Let E_1 denote the event that the sum of the dice is six and F denote the event that the first die equals four. Then

$$P(E_1F) = P(\{4,2\}) = \frac{1}{36}$$

while

$$P(E_1)P(F) = \frac{5}{36}\frac{1}{6} = \frac{5}{216}$$

and hence E_1 and F are not independent. Intuitively, the reason for this is clear for if we are interested in the possibility of throwing a six (with two dice), then we will be quite happy if the first die lands four (or any of the numbers 1, 2, 3, 4, 5) because then we still have a possibility of getting a total of six. On the other hand, if the first die landed six, then we would be unhappy as we would no longer have a chance of getting a total of six. In other words, our chance of getting a total of six depends on the outcome of the first die and hence E_1 and F cannot be independent.

Let E_2 be the event that the sum of the dice equals seven. Is E_2 independent of F? The answer is yes since

$$P(E_2F) = P(\{(4,3)\}) = \frac{1}{36}$$

while

$$P(E_2)P(F) = \frac{1}{6}\frac{1}{6} = \frac{1}{36}$$

We leave it for you to present the intuitive argument why the event that the sum of the dice equals seven is independent of the outcome on the first die. ■

The definition of independence can be extended to more than two events. The events $E_1, E_2, \ldots, E_n$ are said to be independent if for every subset $E_{1'}, E_{2'}, \ldots, E_{r'}$, $r \leqslant n$, of these events

$$P(E_{1'}E_{2'} \cdots E_{r'}) = P(E_{1'})P(E_{2'}) \cdots P(E_{r'})$$

Intuitively, the events $E_1, E_2, \ldots, E_n$ are independent if knowledge of the occurrence of any of these events has no effect on the probability of any other event.

Example 1.10 (Pairwise Independent Events That Are Not Independent) Let a ball be drawn from an urn containing four balls, numbered 1, 2, 3, 4. Let $E = \{1, 2\}$, $F = \{1, 3\}$, $G = \{1, 4\}$. If all four outcomes are assumed equally likely, then

$$\begin{aligned} P(EF) &= P(E)P(F) = \tfrac{1}{4}, \\ P(EG) &= P(E)P(G) = \tfrac{1}{4}, \\ P(FG) &= P(F)P(G) = \tfrac{1}{4} \end{aligned}$$

However,

$$\tfrac{1}{4} = P(EFG) \neq P(E)P(F)P(G)$$

Hence, even though the events E, F, G are pairwise independent, they are not jointly independent. ■

Example 1.11 There are r players, with player i initially having n_i units, $n_i > 0, i = 1, \ldots, r$. At each stage, two of the players are chosen to play a game, with the winner of the game receiving 1 unit from the loser. Any player whose fortune drops to 0 is eliminated, and this continues until a single player has all $n \equiv \sum_{i=1}^{r} n_i$ units, with that player designated as the victor. Assuming that the results of successive games are independent, and that each game is equally likely to be won by either of its two players, find the probability that player i is the victor.

Solution: To begin, suppose that there are n players, with each player initially having 1 unit. Consider player i. Each stage she plays will be equally likely to result in her either winning or losing 1 unit, with the results from each stage being independent. In addition, she will continue to play stages until her fortune becomes either 0 or n. Because this is the same for all players, it follows that each player has the same chance of being the victor. Consequently, each player has player probability $1/n$ of being the victor. Now, suppose these n players are divided into r teams, with team i containing n_i players, $i = 1, \ldots, r$. That is, suppose players $1, \ldots, n_1$ constitute team 1, players $n_1 + 1, \ldots, n_1 + n_2$ constitute team 2 and so on. Then the probability that the victor is a member of team i is n_i/n. But because team i initially has a total fortune of n_i units, $i = 1, \ldots, r$, and each game played by members of different teams results in the fortune of the winner's team increasing by 1 and that of the loser's team decreasing by 1, it is easy to see that the probability that the victor is from team i is exactly the desired probability. Moreover, our argument also shows that the result is true no matter how the choices of the players in each stage are made. ■

Suppose that a sequence of experiments, each of which results in either a "success" or a "failure," is to be performed. Let $E_i, i \geqslant 1$, denote the event that the ith experiment results in a success. If, for all $i_1, i_2, \ldots, i_n$,

$$P(E_{i_1} E_{i_2} \cdots E_{i_n}) = \prod_{j=1}^{n} P(E_{i_j})$$

we say that the sequence of experiments consists of *independent trials*.

1.6 Bayes' Formula

Let E and F be events. We may express E as

$$E = EF \cup EF^c$$

because in order for a point to be in E, it must either be in both E and F, or it must be in E and not in F. Since EF and EF^c are mutually exclusive, we have that

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)(1 - P(F)) \end{aligned} \tag{1.7}$$

Equation (1.7) states that the probability of the event E is a weighted average of the conditional probability of E given that F has occurred and the conditional probability of E given that F has not occurred, each conditional probability being given as much weight as the event on which it is conditioned has of occurring.

Example 1.12 Consider two urns. The first contains two white and seven black balls, and the second contains five white and six black balls. We flip a fair coin and then draw a ball from the first urn or the second urn depending on whether the outcome was heads or tails. What is the conditional probability that the outcome of the toss was heads given that a white ball was selected?

Solution: Let W be the event that a white ball is drawn, and let H be the event that the coin comes up heads. The desired probability $P(H|W)$ may be calculated as follows:

$$\begin{aligned} P(H|W) &= \frac{P(HW)}{P(W)} = \frac{P(W|H)P(H)}{P(W)} \\ &= \frac{P(W|H)P(H)}{P(W|H)P(H) + P(W|H^c)P(H^c)} \\ &= \frac{\frac{2}{9}\frac{1}{2}}{\frac{2}{9}\frac{1}{2} + \frac{5}{11}\frac{1}{2}} = \frac{22}{67} \end{aligned}$$ ■

Example 1.13 In answering a question on a multiple-choice test a student either knows the answer or guesses. Let p be the probability that she knows the answer and $1 - p$ the probability that she guesses. Assume that a student who guesses at the answer will be correct with probability $1/m$, where m is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question given that she answered it correctly?

Solution: Let C and K denote respectively the event that the student answers the question correctly and the event that she actually knows the answer. Now

$$\begin{aligned} P(K|C) &= \frac{P(KC)}{P(C)} = \frac{P(C|K)P(K)}{P(C|K)P(K) + P(C|K^c)P(K^c)} \\ &= \frac{p}{p + (1/m)(1-p)} \\ &= \frac{mp}{1 + (m-1)p} \end{aligned}$$

Thus, for example, if $m = 5, p = \frac{1}{2}$, then the probability that a student knew the answer to a question she correctly answered is $\frac{5}{6}$. ■

Example 1.14 A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability 0.01, the test result will imply he has the disease.) If 0.5 percent of the population actually has the disease, what is the probability a person has the disease given that his test result is positive?

Random Variables

2.1 Random Variables

It frequently occurs that in performing an experiment we are mainly interested in some functions of the outcome as opposed to the outcome itself. For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the actual outcome. That is, we may be interested in knowing that the sum is seven and not be concerned over whether the actual outcome was (1, 6) or (2, 5) or (3, 4) or (4, 3) or (5, 2) or (6, 1). These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as *random variables*.

Since the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable.

Example 2.1 Letting X denote the random variable that is defined as the sum of two fair dice; then

$$
\begin{aligned}
&P\{X=2\}=P\{(1,1)\}=\tfrac{1}{36},\\
&P\{X=3\}=P\{(1,2),(2,1)\}=\tfrac{2}{36},\\
&P\{X=4\}=P\{(1,3),(2,2),(3,1)\}=\tfrac{3}{36},\\
&P\{X=5\}=P\{(1,4),(2,3),(3,2),(4,1)\}=\tfrac{4}{36},\\
&P\{X=6\}=P\{(1,5),(2,4),(3,3),(4,2),(5,1)\}=\tfrac{5}{36},
\end{aligned}
$$

Introduction to Probability Models, ISBN: 9780123756862
Copyright © 2010 by Elsevier, Inc. All rights reserved.

$$
\begin{aligned}
P\{X=7\} &= P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = \tfrac{6}{36},\\
P\{X=8\} &= P\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = \tfrac{5}{36},\\
P\{X=9\} &= P\{(3,6),(4,5),(5,4),(6,3)\} = \tfrac{4}{36},\\
P\{X=10\} &= P\{(4,6),(5,5),(6,4)\} = \tfrac{3}{36},\\
P\{X=11\} &= P\{(5,6),(6,5)\} = \tfrac{2}{36},\\
P\{X=12\} &= P\{(6,6)\} = \tfrac{1}{36} \qquad (2.1)
\end{aligned}
$$

In other words, the random variable X can take on any integral value between two and twelve, and the probability that it takes on each value is given by Equation (2.1). Since X must take on one of the values two through twelve, we must have

$$1 = P\left\{\bigcup_{i=2}^{12}\{X=n\}\right\} = \sum_{n=2}^{12} P\{X=n\}$$

which may be checked from Equation (2.1). ■

Example 2.2 For a second example, suppose that our experiment consists of tossing two fair coins. Letting Y denote the number of heads appearing, then Y is a random variable taking on one of the values 0, 1, 2 with respective probabilities

$$
\begin{aligned}
P\{Y=0\} &= P\{(T,T)\} = \tfrac{1}{4},\\
P\{Y=1\} &= P\{(T,H),(H,T)\} = \tfrac{2}{4},\\
P\{Y=2\} &= P\{(H,H)\} = \tfrac{1}{4}
\end{aligned}
$$

Of course, $P\{Y=0\} + P\{Y=1\} + P\{Y=2\} = 1$. ■

Example 2.3 Suppose that we toss a coin having a probability p of coming up heads, until the first head appears. Letting N denote the number of flips required, then assuming that the outcome of successive flips are independent, N is a random variable taking on one of the values $1, 2, 3, \dots,$ with respective probabilities

$$
\begin{aligned}
P\{N=1\} &= P\{H\} = p,\\
P\{N=2\} &= P\{(T,H)\} = (1-p)p,\\
P\{N=3\} &= P\{(T,T,H)\} = (1-p)^2p,\\
&\ \ \vdots\\
P\{N=n\} &= P\{(\underbrace{T,T,\dots,T}_{n-1},H)\} = (1-p)^{n-1}p, \qquad n \ge 1
\end{aligned}
$$

(i) $F(b)$ is a nondecreasing function of b,
(ii) $\lim_{b\to\infty} F(b) = F(\infty) = 1$,
(iii) $\lim_{b\to-\infty} F(b) = F(-\infty) = 0$.

Property (i) follows since for $a < b$ the event $\{X \le a\}$ is contained in the event $\{X \le b\}$, and so it must have a smaller probability. Properties (ii) and (iii) follow since X must take on some finite value.

All probability questions about X can be answered in terms of the cdf $F(\cdot)$. For example,

$$P\{a < X \le b\} = F(b) - F(a) \qquad \text{for all } a < b$$

This follows since we may calculate $P\{a<X\le b\}$ by first computing the probability that $X \le b$ (that is, $F(b)$) and then subtracting from this the probability that $X \le a$ (that is, $F(a)$).

If we desire the probability that X is strictly smaller than b, we may calculate this probability by

$$\begin{aligned} P\{X < b\} &= \lim_{h\to 0^+} P\{X \le b - h\} \\ &= \lim_{h\to 0^+} F(b-h) \end{aligned}$$

where $\lim_{h\to 0^+}$ means that we are taking the limit as h decreases to 0. Note that $P\{X < b\}$ does not necessarily equal $F(b)$ since $F(b)$ also includes the probability that X equals b.

2.2 Discrete Random Variables

As was previously mentioned, a random variable that can take on at most a countable number of possible values is said to be *discrete*. For a discrete random variable X, we define the *probability mass function* $p(a)$ of X by

$$p(a) = P\{X = a\}$$

The probability mass function $p(a)$ is positive for at most a countable number of values of a. That is, if X must assume one of the values $x_1, x_2, \ldots$, then

$$\begin{aligned} p(x_i) &> 0, \qquad i = 1, 2, \ldots \\ p(x) &= 0, \qquad \text{all other values of } x \end{aligned}$$

Since X must take on one of the values x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

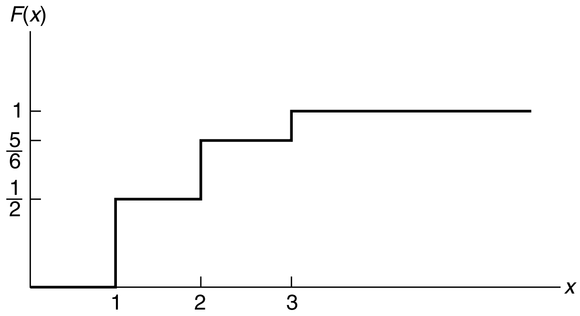

Figure 2.1 Graph of $F(x)$.

The cumulative distribution function F can be expressed in terms of $p(a)$ by

$$F(a) = \sum_{\text{all } x_i \le a} p(x_i)$$

For instance, suppose X has a probability mass function given by

$$p(1) = \tfrac{1}{2}, \qquad p(2) = \tfrac{1}{3}, \qquad p(3) = \tfrac{1}{6}$$

then, the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0, & a < 1 \\ \frac{1}{2}, & 1 \le a < 2 \\ \frac{5}{6}, & 2 \le a < 3 \\ 1, & 3 \le a \end{cases}$$

This is graphically presented in Figure 2.1.

Discrete random variables are often classified according to their probability mass functions. We now consider some of these random variables.

2.2.1 The Bernoulli Random Variable

Suppose that a trial, or an experiment, whose outcome can be classified as either a "success" or as a "failure" is performed. If we let X equal 1 if the outcome is a success and 0 if it is a failure, then the probability mass function of X is given by

$$\begin{aligned} p(0) &= P\{X = 0\} = 1 - p, \\ p(1) &= P\{X = 1\} = p \end{aligned} \tag{2.2}$$

where $p,\ 0 \le p \le 1$, is the probability that the trial is a "success."

A random variable X is said to be a *Bernoulli* random variable if its probability mass function is given by Equation (2.2) for some $p \in (0, 1)$.

2.2.2 The Binomial Random Variable

Suppose that n independent trials, each of which results in a "success" with probability p and in a "failure" with probability $1 - p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a *binomial* random variable with parameters (n, p).

The probability mass function of a binomial random variable having parameters (n, p) is given by

$$p(i) = \binom{n}{i} p^i (1 - p)^{n-i}, \qquad i = 0, 1, \ldots, n \tag{2.3}$$

where

$$\binom{n}{i} = \frac{n!}{(n - i)!\, i!}$$

equals the number of different groups of i objects that can be chosen from a set of n objects. The validity of Equation (2.3) may be verified by first noting that the probability of any particular sequence of the n outcomes containing i successes and $n - i$ failures is, by the assumed independence of trials, $p^i(1 - p)^{n-i}$. Equation (2.3) then follows since there are $\binom{n}{i}$ different sequences of the n outcomes leading to i successes and $n-i$ failures. For instance, if $n = 3, i = 2$, then there are $\binom{3}{2} = 3$ ways in which the three trials can result in two successes. Namely, any one of the three outcomes (s, s, f), (s, f, s), (f, s, s), where the outcome (s, s, f) means that the first two trials are successes and the third a failure. Since each of the three outcomes (s, s, f), (s, f, s), (f, s, s) has a probability $p^2(1 - p)$ of occurring the desired probability is thus $\binom{3}{2} p^2(1 - p)$.

Note that, by the binomial theorem, the probabilities sum to one, that is,

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1 - p)^{n-i} = \big(p + (1 - p)\big)^n = 1$$

Example 2.6 Four fair coins are flipped. If the outcomes are assumed independent, what is the probability that two heads and two tails are obtained?

Solution: Letting X equal the number of heads ("successes") that appear, then X is a binomial random variable with parameters $(n = 4,\ p = \frac{1}{2})$. Hence, by Equation (2.3),

$$P\{X = 2\} = \binom{4}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{3}{8}$$

■

Example 2.11 If the number of accidents occurring on a highway each day is a Poisson random variable with parameter $\lambda = 3$, what is the probability that no accidents occur today?

Solution:

$$P\{X = 0\} = e^{-3} \approx 0.05 \qquad \blacksquare$$

Example 2.12 Consider an experiment that consists of counting the number of α-particles given off in a one-second interval by one gram of radioactive material. If we know from past experience that, on the average, 3.2 such α-particles are given off, what is a good approximation to the probability that no more than two α-particles will appear?

Solution: If we think of the gram of radioactive material as consisting of a large number n of atoms each of which has probability $3.2/n$ of disintegrating and sending off an α-particle during the second considered, then we see that, to a very close approximation, the number of α-particles given off will be a Poisson random variable with parameter $\lambda = 3.2$. Hence the desired probability is

$$P\{X \le 2\} = e^{-3.2} + 3.2e^{-3.2} + \frac{(3.2)^2}{2}e^{-3.2} \approx 0.382 \qquad \blacksquare$$

2.3 Continuous Random Variables

In this section, we shall concern ourselves with random variables whose set of possible values is uncountable. Let X be such a random variable. We say that X is a *continuous* random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x)\,dx \tag{2.6}$$

The function $f(x)$ is called the *probability density function* of the random variable X.

In words, Equation (2.6) states that the probability that X will be in B may be obtained by integrating the probability density function over the set B. Since X must assume some value, $f(x)$ must satisfy

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\,dx$$

All probability statements about X can be answered in terms of $f(x)$. For instance, letting $B = [a, b]$, we obtain from Equation (2.6) that

$$P\{a \leq X \leq b\} = \int_a^b f(x)\, dx \tag{2.7}$$

If we let $a = b$ in the preceding, then

$$P\{X = a\} = \int_a^a f(x)\, dx = 0$$

In words, this equation states that the probability that a continuous random variable will assume any *particular* value is zero.

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)\, dx$$

Differentiating both sides of the preceding yields

$$\frac{d}{da} F(a) = f(a)$$

That is, the density is the derivative of the cumulative distribution function. A somewhat more intuitive interpretation of the density function may be obtained from Equation (2.7) as follows:

$$P\left\{a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}\right\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)\, dx \approx \varepsilon f(a)$$

when ε is small. In other words, the probability that X will be contained in an interval of length ε around the point a is approximately $\varepsilon f(a)$. From this, we see that $f(a)$ is a measure of how likely it is that the random variable will be near a.

There are several important continuous random variables that appear frequently in probability theory. The remainder of this section is devoted to a study of certain of these random variables.

2.3.1 The Uniform Random Variable

A random variable is said to be *uniformly distributed* over the interval $(0, 1)$ if its probability density function is given by

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Note that the preceding is a density function since $f(x) \geq 0$ and

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 dx = 1$$

Since $f(x) > 0$ only when $x \in (0, 1)$, it follows that X must assume a value in $(0, 1)$. Also, since $f(x)$ is constant for $x \in (0, 1)$, X is just as likely to be "near" any value in $(0, 1)$ as any other value. To check this, note that, for any $0 < a < b < 1$,

$$P\{a \leq X \leq b\} = \int_a^b f(x)\,dx = b - a$$

In other words, the probability that X is in any particular subinterval of $(0, 1)$ equals the length of that subinterval.

In general, we say that X is a uniform random variable on the interval (α, β) if its probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \text{if } \alpha < x < \beta \\ 0, & \text{otherwise} \end{cases} \tag{2.8}$$

Example 2.13 Calculate the cumulative distribution function of a random variable uniformly distributed over (α, β).

Solution: Since $F(a) = \int_{-\infty}^{a} f(x)\,dx$, we obtain from Equation (2.8) that

$$F(a) = \begin{cases} 0, & a \leq \alpha \\ \dfrac{a - \alpha}{\beta - \alpha}, & \alpha < a < \beta \\ 1, & a \geq \beta \end{cases}$$

■

Example 2.14 If X is uniformly distributed over $(0, 10)$, calculate the probability that (a) $X < 3$, (b) $X > 7$, (c) $1 < X < 6$.

Solution:

$$P\{X < 3\} = \frac{\int_0^3 dx}{10} = \frac{3}{10},$$

$$P\{X > 7\} = \frac{\int_7^{10} dx}{10} = \frac{3}{10},$$

$$P\{1 < X < 6\} = \frac{\int_1^6 dx}{10} = \frac{1}{2}$$

■

One implication of the preceding result is that if X is normally distributed with parameters μ and σ^2 then $Y = (X-\mu)/\sigma$ is normally distributed with parameters 0 and 1. Such a random variable Y is said to have the *standard* or *unit* normal distribution.

2.4 Expectation of a Random Variable

2.4.1 The Discrete Case

If X is a discrete random variable having a probability mass function $p(x)$, then the *expected value* of X is defined by

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

In other words, the expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes that value. For example, if the probability mass function of X is given by

$$p(1) = \tfrac{1}{2} = p(2)$$

then

$$E[X] = 1(\tfrac{1}{2}) + 2(\tfrac{1}{2}) = \tfrac{3}{2}$$

is just an ordinary average of the two possible values 1 and 2 that X can assume. On the other hand, if

$$p(1) = \tfrac{1}{3}, \qquad p(2) = \tfrac{2}{3}$$

then

$$E[X] = 1(\tfrac{1}{3}) + 2(\tfrac{2}{3}) = \tfrac{5}{3}$$

is a weighted average of the two possible values 1 and 2 where the value 2 is given twice as much weight as the value 1 since $p(2) = 2p(1)$.

Example 2.15 Find $E[X]$ where X is the outcome when we roll a fair die.

Solution: Since $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$, we obtain

$$E[X] = 1(\tfrac{1}{6}) + 2(\tfrac{1}{6}) + 3(\tfrac{1}{6}) + 4(\tfrac{1}{6}) + 5(\tfrac{1}{6}) + 6(\tfrac{1}{6}) = \tfrac{7}{2} \qquad \blacksquare$$

Example 2.16 (Expectation of a Bernoulli Random Variable) Calculate $E[X]$ when X is a Bernoulli random variable with parameter p.

Writing x as $(x-\mu)+\mu$ yields

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)e^{-(x-\mu)^2/2\sigma^2}\,dx + \mu\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}e^{-(x-\mu)^2/2\sigma^2}\,dx$$

Letting $y = x-\mu$ leads to

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}ye^{-y^2/2\sigma^2}\,dy + \mu\int_{-\infty}^{\infty}f(x)\,dx$$

where $f(x)$ is the normal density. By symmetry, the first integral must be 0, and so

$$E[X] = \mu\int_{-\infty}^{\infty}f(x)\,dx = \mu \qquad \blacksquare$$

2.4.3 Expectation of a Function of a Random Variable

Suppose now that we are given a random variable X and its probability distribution (that is, its probability mass function in the discrete case or its probability density function in the continuous case). Suppose also that we are interested in calculating not the expected value of X, but the expected value of some function of X, say, $g(X)$. How do we go about doing this? One way is as follows. Since $g(X)$ is itself a random variable, it must have a probability distribution, which should be computable from a knowledge of the distribution of X. Once we have obtained the distribution of $g(X)$, we can then compute $E[g(X)]$ by the definition of the expectation.

Example 2.23 Suppose X has the following probability mass function:

$$p(0) = 0.2, \qquad p(1) = 0.5, \qquad p(2) = 0.3$$

Calculate $E[X^2]$.

Solution: Letting $Y = X^2$, we have that Y is a random variable that can take on one of the values $0^2, 1^2, 2^2$ with respective probabilities

$$p_Y(0) = P\{Y = 0^2\} = 0.2,$$
$$p_Y(1) = P\{Y = 1^2\} = 0.5,$$
$$p_Y(4) = P\{Y = 2^2\} = 0.3$$

Hence,

$$E[X^2] = E[Y] = 0(0.2) + 1(0.5) + 4(0.3) = 1.7$$

Note that

$$1.7 = E[X^2] \neq (E[X])^2 = 1.21 \qquad \blacksquare$$

Example 2.24 Let X be uniformly distributed over $(0, 1)$. Calculate $E[X^3]$.

Solution: Letting $Y = X^3$, we calculate the distribution of Y as follows. For $0 \leq a \leq 1$,

$$\begin{aligned} F_Y(a) &= P\{Y \leq a\} \\ &= P\{X^3 \leq a\} \\ &= P\{X \leq a^{1/3}\} \\ &= a^{1/3} \end{aligned}$$

where the last equality follows since X is uniformly distributed over $(0, 1)$. By differentiating $F_Y(a)$, we obtain the density of Y, namely,

$$f_Y(a) = \tfrac{1}{3}a^{-2/3}, \qquad 0 \leq a \leq 1$$

Hence,

$$\begin{aligned} E[X^3] = E[Y] &= \int_{-\infty}^{\infty} a f_Y(a)\, da \\ &= \int_0^1 a \tfrac{1}{3} a^{-2/3}\, da \\ &= \tfrac{1}{3}\int_0^1 a^{1/3}\, da \\ &= \tfrac{1}{3}\tfrac{3}{4} a^{4/3}\Big|_0^1 \\ &= \tfrac{1}{4} \end{aligned} \qquad \blacksquare$$

While the foregoing procedure will, in theory, always enable us to compute the expectation of any function of X from a knowledge of the distribution of X, there is, fortunately, an easier way to do this. The following proposition shows how we can calculate the expectation of $g(X)$ without first determining its distribution.

Proposition 2.1 (a) If X is a discrete random variable with probability mass function $p(x)$, then for any real-valued function g,

$$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$$

of X. By Proposition 2.1, we note that

$$E[X^n] = \begin{cases} \displaystyle\sum_{x:p(x)>0} x^n p(x), & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} x^n f(x)\,dx, & \text{if } X \text{ is continuous} \end{cases}$$

Another quantity of interest is the variance of a random variable X, denoted by $\mathrm{Var}(X)$, which is defined by

$$\mathrm{Var}(X) = E\big[(X - E[X])^2\big]$$

Thus, the variance of X measures the expected square of the deviation of X from its expected value.

Example 2.27 (Variance of the Normal Random Variable) Let X be normally distributed with parameters μ and σ^2. Find $\mathrm{Var}(X)$.

Solution: Recalling (see Example 2.22) that $E[X] = \mu$, we have that

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2}\,dx\end{aligned}$$

Substituting $y = (x-\mu)/\sigma$ yields

$$\mathrm{Var}(X) = \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-y^2/2}\,dy$$

Integrating by parts ($u = y,\ dv = ye^{-y^2/2}dy$) gives

$$\begin{aligned}\mathrm{Var}(X) &= \frac{\sigma^2}{\sqrt{2\pi}}\left(-ye^{-y^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2}\,dy\right) \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}\,dy \\ &= \sigma^2\end{aligned}$$

Another derivation of $\mathrm{Var}(X)$ will be given in Example 2.42. ■

Suppose that X is continuous with density f, and let $E[X] = \mu$. Then,

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2]\end{aligned}$$

$$= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)\, dx$$
$$= \int_{-\infty}^{\infty} x^2 f(x)\, dx - 2\mu \int_{-\infty}^{\infty} x f(x)\, dx + \mu^2 \int_{-\infty}^{\infty} f(x)\, dx$$
$$= E[X^2] - 2\mu\mu + \mu^2$$
$$= E[X^2] - \mu^2$$

A similar proof holds in the discrete case, and so we obtain the useful identity

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

Example 2.28 Calculate Var(X) when X represents the outcome when a fair die is rolled.

Solution: As previously noted in Example 2.15, $E[X] = \frac{7}{2}$. Also,

$$E[X^2] = 1\left(\tfrac{1}{6}\right) + 2^2\left(\tfrac{1}{6}\right) + 3^2\left(\tfrac{1}{6}\right) + 4^2\left(\tfrac{1}{6}\right) + 5^2\left(\tfrac{1}{6}\right) + 6^2\left(\tfrac{1}{6}\right) = \left(\tfrac{1}{6}\right)(91)$$

Hence,

$$\mathrm{Var}(X) = \tfrac{91}{6} - \left(\tfrac{7}{2}\right)^2 = \tfrac{35}{12}$$ ■

2.5 Jointly Distributed Random Variables

2.5.1 Joint Distribution Functions

Thus far, we have concerned ourselves with the probability distribution of a single random variable. However, we are often interested in probability statements concerning two or more random variables. To deal with such probabilities, we define, for any two random variables X and Y, the *joint cumulative probability distribution function* of X and Y by

$$F(a, b) = P\{X \le a, Y \le b\}, \qquad -\infty < a, b < \infty$$

The distribution of X can be obtained from the joint distribution of X and Y as follows:

$$\begin{aligned} F_X(a) &= P\{X \le a\} \\ &= P\{X \le a,\ Y < \infty\} \\ &= F(a, \infty) \end{aligned}$$

Similarly, the cumulative distribution function of Y is given by

$$F_Y(b) = P\{Y \le b\} = F(\infty,\ b)$$

In the case where X and Y are both discrete random variables, it is convenient to define the *joint probability mass function* of X and Y by

$$p(x, y) = P\{X = x,\ Y = y\}$$

The probability mass function of X may be obtained from $p(x, y)$ by

$$p_X(x) = \sum_{y:p(x,y)>0} p(x, y)$$

Similarly,

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x, y)$$

We say that X and Y are *jointly continuous* if there exists a function $f(x, y)$, defined for all real x and y, having the property that for all sets A and B of real numbers

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y)\, dx\, dy$$

The function $f(x, y)$ is called the *joint probability density function* of X and Y. The probability density of X can be obtained from a knowledge of $f(x, y)$ by the following reasoning:

$$\begin{aligned} P\{X \in A\} &= P\{X \in A,\ Y \in (-\infty, \infty)\} \\ &= \int_{-\infty}^{\infty} \int_A f(x, y)\, dx\, dy \\ &= \int_A f_X(x)\, dx \end{aligned}$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

is thus the probability density function of X. Similarly, the probability density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

Because

$$F(a,b) = P(X \le a, Y \le b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y)dy\,dx$$

differentiation yields

$$\frac{d^2}{da\,db}F(a,b) = f(a,b)$$

Thus, as in the single variable case, differentiating the probability distribution function gives the probability density function.

A variation of Proposition 2.1 states that if X and Y are random variables and g is a function of two variables, then

$$\begin{aligned} E[g(X,Y)] &= \sum_y \sum_x g(x,y)p(x,y) && \text{in the discrete case} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy && \text{in the continuous case} \end{aligned}$$

For example, if $g(X,Y) = X + Y$, then, in the continuous case,

$$\begin{aligned} E[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)f(x,y)\,dx\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,dx\,dy \\ &= E[X] + E[Y] \end{aligned}$$

where the first integral is evaluated by using the variation of Proposition 2.1 with $g(x,y) = x$, and the second with $g(x,y) = y$.

The same result holds in the discrete case and, combined with the corollary in Section 2.4.3, yields that for any constants a, b

$$E[aX + bY] = aE[X] + bE[Y] \tag{2.10}$$

Joint probability distributions may also be defined for n random variables. The details are exactly the same as when $n = 2$ and are left as an exercise. The corresponding result to Equation (2.10) states that if $X_1, X_2, \dots, X_n$ are n random variables, then for any n constants $a_1, a_2, \dots, a_n$,

$$E[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1E[X_1] + a_2E[X_2] + \cdots + a_nE[X_n] \tag{2.11}$$

and so

$$E[X_i] = 1P\{X_i = 1\} + 0P\{X_i = 0\} = \frac{1}{N}$$

Hence, from Equation (2.11) we obtain

$$E[X] = E[X_1] + \cdots + E[X_N] = \left(\frac{1}{N}\right)N = 1$$

Hence, no matter how many people are at the party, on the average exactly one of the men will select his own hat. ■

Example 2.32 Suppose there are 25 different types of coupons and suppose that each time one obtains a coupon, it is equally likely to be any one of the 25 types. Compute the expected number of different types that are contained in a set of 10 coupons.

Solution: Let X denote the number of different types in the set of 10 coupons. We compute $E[X]$ by using the representation

$$X = X_1 + \cdots + X_{25}$$

where

$$X_i = \begin{cases} 1, & \text{if at least one type } i \text{ coupon is in the set of 10} \\ 0, & \text{otherwise} \end{cases}$$

Now,

$$\begin{aligned} E[X_i] &= P\{X_i = 1\} \\ &= P\{\text{at least one type } i \text{ coupon is in the set of 10}\} \\ &= 1 - P\{\text{no type } i \text{ coupons are in the set of 10}\} \\ &= 1 - \left(\tfrac{24}{25}\right)^{10} \end{aligned}$$

when the last equality follows since each of the 10 coupons will (independently) not be a type i with probability $\frac{24}{25}$. Hence,

$$E[X] = E[X_1] + \cdots + E[X_{25}] = 25\left[1 - \left(\tfrac{24}{25}\right)^{10}\right]$$ ■

2.5.2 Independent Random Variables

The random variables X and Y are said to be *independent* if, for all a, b,

$$P\{X \le a,\ Y \le b\} = P\{X \le a\}P\{Y \le b\} \tag{2.12}$$

In other words, X and Y are independent if, for all a and b, the events $E_a = \{X \leq a\}$ and $F_b = \{Y \leq b\}$ are independent.

In terms of the joint distribution function F of X and Y, we have that X and Y are independent if

$$F(a, b) = F_X(a)F_Y(b) \qquad \text{for all } a, b$$

When X and Y are discrete, the condition of independence reduces to

$$p(x, y) = p_X(x)p_Y(y) \tag{2.13}$$

while if X and Y are jointly continuous, independence reduces to

$$f(x, y) = f_X(x)f_Y(y) \tag{2.14}$$

To prove this statement, consider first the discrete version, and suppose that the joint probability mass function $p(x, y)$ satisfies Equation (2.13). Then

$$\begin{aligned} P\{X \leq a, Y \leq b\} &= \sum_{y \leq b} \sum_{x \leq a} p(x, y) \\ &= \sum_{y \leq b} \sum_{x \leq a} p_X(x)p_Y(y) \\ &= \sum_{y \leq b} p_Y(y) \sum_{x \leq a} p_X(x) \\ &= P\{Y \leq b\}P\{X \leq a\} \end{aligned}$$

and so X and Y are independent. That Equation (2.14) implies independence in the continuous case is proven in the same manner and is left as an exercise.

An important result concerning independence is the following.

Proposition 2.3 If X and Y are independent, then for any functions h and g

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Proof. Suppose that X and Y are jointly continuous. Then

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} h(y)f_Y(y)\, dy \int_{-\infty}^{\infty} g(x)f_X(x)\, dx \\ &= E[h(Y)]E[g(X)] \end{aligned}$$

The proof in the discrete case is similar. ∎