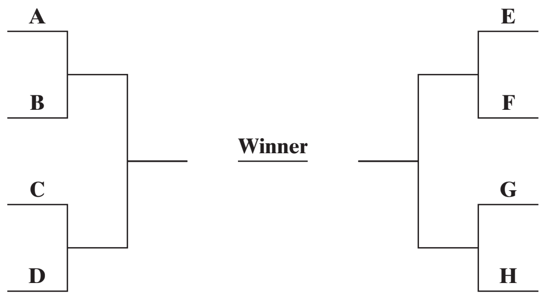

For example, eight students (called A–H) set up a tournament among themselves. The top-listed student in each bracket calls heads or tails when his or her opponent flips a coin. If the call is correct, the student moves on to the next bracket.

(a) How many coin flips are required to determine the tournament winner?

(b) What is the probability that you can predict all of the winners?

(c) In NCAA Division I basketball, after the "play-in" games, 64 teams participate in a single-elimination tournament to determine the national champion. Considering only the remaining 64 teams, how many games are required to determine the national champion?

(d) Assume that for any given game, either team has an equal chance of winning. (That is probably not true.) On page 43 of the March 22, 1999, issue, *Time* claimed that the "mathematical odds of predicting all 63 NCAA games correctly is 1 in 75 million." Do you agree with this statement? If not, why not?

1.4-19. Extend Example 1.4-6 to an n-sided die. That is, suppose that a fair n-sided die is rolled n independent times. A match occurs if side i is observed on the ith trial, $i = 1, 2, \ldots, n$.

(a) Show that the probability of at least one match is

$$1 - \left(\frac{n-1}{n}\right)^n = 1 - \left(1 - \frac{1}{n}\right)^n.$$

(b) Find the limit of this probability as n increases without bound.

1.4-20. Hunters A and B shoot at a target with probabilities of p_1 and p_2, respectively. Assuming independence, can p_1 and p_2 be selected so that $P(\text{zero hits}) = P(\text{one hit}) = P(\text{two hits})$?

1.5 BAYES' THEOREM

We begin this section by illustrating Bayes' theorem with an example.

Example 1.5-1 Bowl B_1 contains two red and four white chips, bowl B_2 contains one red and two white chips, and bowl B_3 contains five red and four white chips. Say that the probabilities for selecting the bowls are not the same but are given by $P(B_1) = 1/3$, $P(B_2) = 1/6$, and $P(B_3) = 1/2$, where B_1, B_2, and B_3 are the events that bowls B_1, B_2, and B_3 are respectively chosen. The experiment consists of selecting a bowl with these probabilities and then drawing a chip at random from that bowl. Let us compute the probability of event R, drawing a red chip—say, $P(R)$. Note that $P(R)$ is dependent first of all on which bowl is selected and then on the probability of drawing a red chip from the selected bowl. That is, the event R is the union of the mutually exclusive events $B_1 \cap R$, $B_2 \cap R$, and $B_3 \cap R$. Thus,

$$\begin{aligned} P(R) &= P(B_1 \cap R) + P(B_2 \cap R) + P(B_3 \cap R) \\ &= P(B_1)P(R \mid B_1) + P(B_2)P(R \mid B_2) + P(B_3)P(R \mid B_3) \\ &= \frac{1}{3} \cdot \frac{2}{6} + \frac{1}{6} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{5}{9} = \frac{4}{9}. \end{aligned}$$

Suppose now that the outcome of the experiment is a red chip, but we do not know from which bowl it was drawn. Accordingly, we compute the conditional probability that the chip was drawn from bowl B_1, namely, $P(B_1 \mid R)$. From the definition of conditional probability and the preceding result, we have

$$\begin{aligned}P(B_1 \mid R) &= \frac{P(B_1 \cap R)}{P(R)} \\ &= \frac{P(B_1)P(R \mid B_1)}{P(B_1)P(R \mid B_1) + P(B_2)P(R \mid B_2) + P(B_3)P(R \mid B_3)} \\ &= \frac{(1/3)(2/6)}{(1/3)(2/6) + (1/6)(1/3) + (1/2)(5/9)} = \frac{2}{8}.\end{aligned}$$

Similarly,

$$P(B_2 \mid R) = \frac{P(B_2 \cap R)}{P(R)} = \frac{(1/6)(1/3)}{4/9} = \frac{1}{8}$$

and

$$P(B_3 \mid R) = \frac{P(B_3 \cap R)}{P(R)} = \frac{(1/2)(5/9)}{4/9} = \frac{5}{8}.$$

Note that the conditional probabilities $P(B_1 \mid R)$, $P(B_2 \mid R)$, and $P(B_3 \mid R)$ have changed from the original probabilities $P(B_1)$, $P(B_2)$, and $P(B_3)$ in a way that agrees with your intuition. Once the red chip has been observed, the probability concerning B_3 seems more favorable than originally because B_3 has a larger percentage of red chips than do B_1 and B_2. The conditional probabilities of B_1 and B_2 decrease from their original ones once the red chip is observed. Frequently, the original probabilities are called *prior probabilities* and the conditional probabilities are the *posterior probabilities*. ■

We generalize the result of Example 1.5-1. Let $B_1, B_2, \ldots, B_m$ constitute a *partition* of the sample space S. That is,

$$S = B_1 \cup B_2 \cup \cdots \cup B_m \text{ and } B_i \cap B_j = \emptyset, i \neq j.$$

Of course, the events $B_1, B_2, \ldots, B_m$ are mutually exclusive and exhaustive (since the union of the disjoint sets equals the sample space S). Furthermore, suppose the **prior probability** of the event B_i is positive; that is, $P(B_i) > 0$, $i = 1, \ldots, m$. If A is an event, then A is the union of m mutually exclusive events, namely,

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_m \cap A).$$

Thus,

$$\begin{aligned}P(A) &= \sum_{i=1}^{m} P(B_i \cap A) \\ &= \sum_{i=1}^{m} P(B_i)P(A \mid B_i), \end{aligned} \qquad (1.5\text{-}1)$$

which is sometimes called the **law of total probability**. If $P(A) > 0$, then

$$P(B_k \mid A) = \frac{P(B_k \cap A)}{P(A)}, \qquad k = 1, 2, \ldots, m. \qquad (1.5\text{-}2)$$

Using Equation 1.5-1 and replacing $P(A)$ in Equation 1.5-2, we have **Bayes' theorem**:

$$P(B_k \mid A) = \frac{P(B_k)P(A \mid B_k)}{\sum_{i=1}^{m} P(B_i)P(A \mid B_i)}, \qquad k = 1, 2, \ldots, m.$$

The conditional probability $P(B_k \mid A)$ is often called the **posterior probability** of B_k. The next example illustrates one application of Bayes' theorem.

Example 1.5-2 In a certain factory, machines I, II, and III are all producing springs of the same length. Of their production, machines I, II, and III respectively produce 2%, 1%, and 3% defective springs. Of the total production of springs in the factory, machine I produces 35%, machine II produces 25%, and machine III produces 40%. If one spring is selected at random from the total springs produced in a day, by the law of total probability, $P(D)$ equals, in an obvious notation,

$$\begin{aligned} P(D) &= P(I)P(D \mid I) + P(II)P(D \mid II) + P(III)P(D \mid III) \\ &= \left(\frac{35}{100}\right)\left(\frac{2}{100}\right) + \left(\frac{25}{100}\right)\left(\frac{1}{100}\right) + \left(\frac{40}{100}\right)\left(\frac{3}{100}\right) = \frac{215}{10{,}000}. \end{aligned}$$

If the selected spring is defective, the conditional probability that it was produced by machine III is, by Bayes' formula,

$$P(III \mid D) = \frac{P(III)P(D \mid III)}{P(D)} = \frac{(40/100)(3/100)}{215/10{,}000} = \frac{120}{215}.$$

Note how the posterior probability of III increased from the prior probability of III after the defective spring was observed, because III produces a larger percentage of defectives than do I and II.

Example 1.5-3 A Pap smear is a screening procedure used to detect cervical cancer. For women with this cancer, there are about 16% *false negatives*; that is,

$$P(T^- = \text{test negative} \mid C^+ = \text{cancer}) = 0.16.$$

Thus,

$$P(T^+ = \text{test positive} \mid C^+ = \text{cancer}) = 0.84.$$

For women without cancer, there are about 10% false positives; that is,

$$P(T^+ \mid C^- = \text{not cancer}) = 0.10.$$

Hence,

$$P(T^- \mid C^- = \text{not cancer}) = 0.90.$$

In the United States, there are about 8 women in 100,000 who have this cancer; that is,

$$P(C^+) = 0.00008; \text{ so } P(C^-) = 0.99992.$$

By Bayes' theorem and the law of total probability,

$$P(C^+ \mid T^+) = \frac{P(C^+ \text{ and } T^+)}{P(T^+)}$$
$$= \frac{(0.00008)(0.84)}{(0.00008)(0.84) + (0.99992)(0.10)}$$
$$= \frac{672}{672 + 999{,}920} = 0.000672.$$

What this means is that for every million positive Pap smears, only 672 represent true cases of cervical cancer. This low ratio makes one question the value of the procedure. The reason that it is ineffective is that the percentage of women having that cancer is so small and the error rates of the procedure—namely, 0.16 and 0.10—are so high. On the other hand, the test does give good information in a sense. The posterior probability of cancer, given a positive test, is about eight times the prior probability. ■

Exercises

1.5-1. Bowl B_1 contains two white chips, bowl B_2 contains two red chips, bowl B_3 contains two white and two red chips, and bowl B_4 contains three white chips and one red chip. The probabilities of selecting bowl B_1, B_2, B_3, or B_4 are 1/2, 1/4, 1/8, and 1/8, respectively. A bowl is selected using these probabilities and a chip is then drawn at random. Find

(a) $P(W)$, the probability of drawing a white chip.

(b) $P(B_1 \mid W)$, the conditional probability that bowl B_1 had been selected, given that a white chip was drawn.

1.5-2. Bean seeds from supplier A have an 85% germination rate and those from supplier B have a 75% germination rate. A seed-packaging company purchases 40% of its bean seeds from supplier A and 60% from supplier B and mixes these seeds together.

(a) Find the probability $P(G)$ that a seed selected at random from the mixed seeds will germinate.

(b) Given that a seed germinates, find the probability that the seed was purchased from supplier A.

1.5-3. A doctor is concerned about the relationship between blood pressure and irregular heartbeats. Among her patients, she classifies blood pressures as high, normal, or low and heartbeats as regular or irregular and finds that (a) 16% have high blood pressure; (b) 19% have low blood pressure; (c) 17% have an irregular heartbeat; (d) of those with an irregular heartbeat, 35% have high blood pressure; and (e) of those with normal blood pressure, 11% have an irregular heartbeat. What percentage of her patients have a regular heartbeat and low blood pressure?

1.5-4. Assume that an insurance company knows the following probabilities relating to automobile accidents (where the second column refers to the probability that the policyholder has at least one accident during the annual policy period):

Age of Driver	Probability of Accident	Fraction of Company's Insured Drivers
16–25	0.05	0.10
26–50	0.02	0.55
51–65	0.03	0.20
66–90	0.04	0.15

A randomly selected driver from the company's insured drivers has an accident. What is the conditional probability that the driver is in the 16–25 age group?

1.5-5. At a hospital's emergency room, patients are classified and 20% of them are critical, 30% are serious, and 50% are stable. Of the critical ones, 30% die; of the serious, 10% die; and of the stable, 1% die. Given that a patient dies, what is the conditional probability that the patient was classified as critical?

1.5-6. A life insurance company issues standard, preferred, and ultrapreferred policies. Of the company's policyholders of a certain age, 60% have standard policies and a probability of 0.01 of dying in the next year, 30% have preferred policies and a probability of 0.008 of dying in the next year, and 10% have ultrapreferred policies and a probability of 0.007 of dying in the next year. A policyholder of that age dies in the next year. What are the conditional probabilities of the deceased having had a standard, a preferred, and an ultrapreferred policy?

1.5-7. A chemist wishes to detect an impurity in a certain compound that she is making. There is a test that detects an impurity with probability 0.90; however, this test indicates that an impurity is there when it is not about 5% of the time. The chemist produces compounds with the impurity about 20% of the time; that is, 80% do not have the impurity. A compound is selected at random from the chemist's output. The test indicates that an impurity is present. What is the conditional probability that the compound actually has an impurity?

1.5-8. A store sells four brands of tablets. The least expensive brand, B_1, accounts for 40% of the sales. The other brands (in order of their price) have the following percentages of sales: $B_2, 30\%$; $B_3, 20\%$; and $B_4, 10\%$. The respective probabilities of needing repair during warranty are 0.10 for B_1, 0.05 for B_2, 0.03 for B_3, and 0.02 for B_4. A randomly selected purchaser has a tablet that needs repair under warranty. What are the four conditional probabilities of being brand B_i, $i = 1, 2, 3, 4$?

1.5-9. There is a new diagnostic test for a disease that occurs in about 0.05% of the population. The test is not perfect, but will detect a person with the disease 99% of the time. It will, however, say that a person without the disease has the disease about 3% of the time. A person is selected at random from the population, and the test indicates that this person has the disease. What are the conditional probabilities that

(a) the person has the disease?

(b) the person does not have the disease?

Discuss. HINT: Note that the fraction 0.0005 of diseased persons in the population is much smaller than the error probabilities of 0.01 and 0.03.

1.5-10. Suppose we want to investigate the percentage of abused children in a certain population. To do this, doctors examine some of these children taken at random from that population. However, doctors are not perfect: They sometimes classify an abused child (A^+) as one not abused (D^-) or they classify a nonabused child (A^-) as one that is abused (D^+). Suppose these error rates are $P(D^- \mid A^+) = 0.08$ and $P(D^+ \mid A^-) = 0.05$, respectively; thus, $P(D^+ \mid A^+) = 0.92$ and $P(D^- \mid A^-) = 0.95$ are the probabilities of the correct decisions. Let us pretend that only 2% of all children are abused; that is, $P(A^+) = 0.02$ and $P(A^-) = 0.98$.

(a) Select a child at random. What is the probability that the doctor classifies this child as abused? That is, compute

$$P(D^+) = P(A^+)P(D^+ \mid A^+) + P(A^-)P(D^+ \mid A^-).$$

(b) Compute $P(A^- \mid D^+)$ and $P(A^+ \mid D^+)$.

(c) Compute $P(A^- \mid D^-)$ and $P(A^+ \mid D^-)$.

(d) Are the probabilities in (b) and (c) alarming? This happens because the error rates of 0.08 and 0.05 are high relative to the fraction 0.02 of abused children in the population.

1.5-11. At the beginning of a certain study of a group of persons, 15% were classified as heavy smokers, 30% as light smokers, and 55% as nonsmokers. In the five-year study, it was determined that the death rates of the heavy and light smokers were five and three times that of the nonsmokers, respectively. A randomly selected participant died over the five-year period; calculate the probability that the participant was a nonsmoker.

1.5-12. A test indicates the presence of a particular disease 90% of the time when the disease is present and the presence of the disease 2% of the time when the disease is not present. If 0.5% of the population has the disease, calculate the conditional probability that a person selected at random has the disease if the test indicates the presence of the disease.

1.5-13. A hospital receives two fifths of its flu vaccine from Company A and the remainder from Company B. Each shipment contains a large number of vials of vaccine. From Company A, 3% of the vials are ineffective; from Company B, 2% are ineffective. A hospital tests $n = 25$ randomly selected vials from one shipment and finds that 2 are ineffective. What is the conditional probability that this shipment came from Company A?

1.5-14. Two processes of a company produce rolls of materials: The rolls of Process I are 3% defective and the rolls of Process II are 1% defective. Process I produces 60% of the company's output, Process II 40%. A roll is selected at random from the total output. Given that this roll is defective, what is the conditional probability that it is from Process I?

HISTORICAL COMMENTS Most probabilists would say that the mathematics of probability began when, in 1654, Chevalier de Méré, a French nobleman who liked to gamble, challenged Blaise Pascal to explain a puzzle and a problem created from his observations concerning rolls of dice. Of course, there was gambling well before this, and actually, almost 200 years before this challenge, a Franciscan monk, Luca Paccioli, proposed essentially the same puzzle. Here it is:

> A and B are playing a fair game of balla. They agree to continue until one has six wins. However, the game actually stops when A has won five and B three. How should the stakes be divided?

And over 100 years before de Méré's challenge, a 16th-century doctor, Girolamo Cardano, who was also a gambler, had figured out the answers to many dice problems, but not the one that de Méré proposed. Chevalier de Méré had observed this: If a single fair die is tossed 4 times, the probability of obtaining at least one six was slightly greater than 1/2. However, keeping the same proportions, if a pair of dice is tossed 24 times, the probability of obtaining at least one double-six seemed to be slightly less than 1/2; at least de Méré was losing money betting on it. This is when he approached Blaise Pascal with the challenge. Not wanting to work on the problems alone, Pascal formed a partnership with Pierre de Fermat, a brilliant young mathematician. It was this 1654 correspondence between Pascal and Fermat that started the theory of probability.

Today an average student in probability could solve both problems easily. For the puzzle, note that B could win with six rounds only by winning the next three rounds, which has probability of $(1/2)^3 = 1/8$ because it was a fair game of balla. Thus, A's probability of winning six rounds is $1 - 1/8 = 7/8$, and stakes should be divided seven units to one. For the dice problem, the probability of at least one six in four rolls of a die is

$$1 - \left(\frac{5}{6}\right)^4 = 0.518,$$

while the probability of rolling at least one double-six in 24 rolls of a pair of dice is

$$1 - \left(\frac{35}{36}\right)^{24} = 0.491.$$

It seems amazing to us that de Méré could have observed enough trials of those events to detect the slight difference in those probabilities. However, he won betting on the first but lost by betting on the second.

Incidentally, the solution to the balla puzzle led to a generalization—namely, the binomial distribution—and to the famous Pascal triangle. Of course, Fermat was the great mathematician associated with "Fermat's last theorem."

The Reverend Thomas Bayes, who was born in 1701, was a Nonconformist (a Protestant who rejected most of the rituals of the Church of England). While he published nothing in mathematics when he was alive, two works were published after his death, one of which contained the essence of Bayes' theorem and a very original way of using data to modify prior probabilities to create posterior probabilities. It has had such an influence on modern statistics that many modern statisticians are associated with the neo-Bayesian movement and we devote Sections 6.8 and 6.9 to some of these methods.