and 1 called the quantum efficiency of the device. We can think of $W/h\nu$ as the total number of photons striking the surface and of η as the probability that a single photon ejects an electron that is counted in the external circuit.

The volume to which the light penetrates just beneath the surface contains a huge number n of conduction electrons. The probability p that a single such electron absorbs a photon, makes its way to the surface, and is drawn over to the anode is extremely small. If we term the emission of a photoelectron and its registering in the external circuit a success, the probability that k electrons are counted will be given by a binomial distribution as in (1.25), but because $n >> 1$, $p << 1$, that probability can be approximated by the Poisson formula in (1.27). It is assumed that whether a given electron absorbs a photon and succeeds in being counted is independent of whether any other electron does so, and that the light is not so intense as significantly to deplete the number of available electrons.

EXAMPLE 1.20

Suppose that the expected number of electrons counted in $(0, T)$ is $m = 4.5$. The probability that exactly two electrons are counted is then

$$\Pr(A_2) = \frac{(4.5)^2}{2} e^{-4.5} = 0.1125.$$

The probability that at least six electrons are counted is

$$\Pr(k \geq 6) = 1 - \sum_{k=0}^{5} \frac{(4.5)^k}{k!} e^{-4.5} = 0.2971.$$

A calculator program for the cumulative Poisson distribution

$$P(k; m) = \sum_{r=0}^{k} \frac{m^r}{r!} e^{-m} = \sum_{r=0}^{k} T_r, \qquad T_r = \frac{m^r}{r!} e^{-m}, \tag{1.29}$$

is easily composed. Starting with $T_0 = e^{-m}$, it calculates the terms T_r successively by the recurrent relation

$$T_{r+1} = \left(\frac{m}{r+1}\right) T_r$$

and accumulates their sum until r reaches the upper limit k. You can check your own program by the following table for $m = 2$.

k	0	1	2	3	4	5	6
$P(k; 2)$	0.13534	0.40601	0.67668	0.85712	0.94735	0.98344	0.99547

If m and k are so large that computation of (1.29) consumes too much time, we can approximate that sum by a convenient formula that will be derived in Sec. 4-7.

Here are some other situations in which the Poisson distribution is appropriate. A microprocessor contains, among other features, a large number n of identical circuits on a silicon chip, and the probability p that any one circuit is defective is very small. The number k of defective circuits on the chip has a Poisson distribution (1.27) with mean np. A biologist searches for a certain species of diatom in a bucket of water drawn from the ocean; there are $n >> 1$ diatoms in the sea, but the probability p that a particular one finds itself in the bucket is very small; the Poisson distribution governs the total number of diatoms caught. The number n of cars on the freeway during a certain hour is very large, and the probability p that a particular car would have an accident during that time is very small; the total number of accidents during that hour is described by a Poisson distribution.

The Poisson approximation to the binomial distribution in (1.27) can be derived from Stirling's formula for the factorial,

$$n! \simeq (2\pi)^{1/2} n^{n+1/2} e^{-n}, \qquad n >> 1, \tag{1.30}$$

of which the relative error is of the order of $1/(12n)$ and small even for $n = 10$. Now in (1.27) both n and $n - k$ are huge, and we use Stirling's formula for both $n!$ and $(n-k)!$. Because $p << 1$, $n >> k$, we can put

$$(1-p)^{n-k} = \left(1 - \frac{m}{n}\right)^{n-k} \simeq \left(1 - \frac{m}{n}\right)^{n} \simeq e^{-m}, \qquad m = np,$$

applying the rule

$$\left(1 + \frac{x}{n}\right)^{n} \simeq e^{x}, \qquad n >> |x|. \tag{1.31}$$

Thus we find that

$$\begin{aligned}
\Pr(A_k) &= \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} \\
&\simeq \frac{n^{n+1/2} e^{-n}}{k!\,(n-k)^{n-k+1/2} e^{-(n-k)}} \left(\frac{m}{n}\right)^k e^{-m} \\
&= \frac{m^k}{k!} e^{-m} \left(1 - \frac{k}{n}\right)^{-(n-k+1/2)} e^{-k}.
\end{aligned}$$

Because $n >> k$, we can apply (1.31) to the next-to-last factor, which becomes e^k and cancels the factor e^{-k}, and we obtain (1.27).

1.10 Randomly Selected Experiments and Hypothesis Testing

Of two possible chance experiments E_1 and E_2 that might be performed, the one that is performed is selected by tossing a coin. Let p be the probability that E_1 is selected, and $1 - p$ the probability that E_2 is selected. This procedure could in principle be repeated arbitrarily often, so that now E_1, now E_2 is performed, with relative frequencies p and $1 - p$, respectively. It corresponds to an experiment E

that is called the *sum* of experiments E_1 and E_2 and sometimes designated by $\mathsf{E}_1 + \mathsf{E}_2$, although the notation $p\mathsf{E}_1 + (1-p)\mathsf{E}_2$ is more informative. The outcomes of the *composite experiment* E are all the outcomes of E_1 conjoined with all the outcomes of E_2.

EXAMPLE 1.21

Upon entering a gambling den you are asked to toss a coin to determine whether you shall play blackjack or roulette. The probability of winning at blackjack is 0.36, that of winning at roulette is 0.44. If the coin is fair, the probability of your departing a winner is $\frac{1}{2}(0.36) + \frac{1}{2}(0.44) = 0.4$.

This concept can be extended to selections among any number of chance experiments. If experiment E_k is chosen with probability p_k, $1 \le k \le n$, the "sum" of these experiments can be designated by $\mathsf{E} = p_1\mathsf{E}_1 + p_2\mathsf{E}_2 + \cdots + p_n\mathsf{E}_n$,

$$\sum_{k=1}^{n} p_k = 1.$$

With respect to this composite experiment E, the probability of event A is, by the principle of total probability,

$$\Pr(A) = \sum_{k=1}^{n} p_k \Pr_k(A), \tag{1.32}$$

where $\Pr_k(A)$ is the probability of event A in experiment E_k and can be considered as a conditional probability with respect to E,

$$\Pr_k(A) = \Pr(A \mid \mathsf{E}_k \text{ chosen}). \tag{1.33}$$

If none of the outcomes of A is among those of E_k, that is, if in E_k $A = \varnothing$, this probability equals zero.

An example of such a composite experiment is Example 1.8, in which picking a transistor from the bin and determining whether it operates for 1000 hours can be viewed as selecting experiment E_a, E_b, or E_c with probability 0.45, 0.25, and 0.30, respectively. Experiment E_a is to test the transistors made by manufacturer A, and so on.

This concept of a composite experiment arises principally in what statisticians call *hypothesis testing*. The outcomes of all n experiments $\mathsf{E}_1, \mathsf{E}_2, \ldots, \mathsf{E}_n$ now lie in the same universal set S, but the probabilities of events in S depend on which experiment is performed. Nature or some other inscrutable agent selects the experiment that is actually performed and picks E_k with relative frequency p_k. The proposition "Experiment E_k was selected and performed" is called a *hypothesis* and denoted by H_k, $k = 1, 2, \ldots, n$. It specifies an event in the composite experiment $\mathsf{E} = p_1\mathsf{E}_1 + p_2\mathsf{E}_2 + \cdots + p_n\mathsf{E}_n$.

The outcomes of the composite experiment $\mathsf{E} = p_1\mathsf{E}_1 + p_2\mathsf{E}_2 + \cdots + p_n\mathsf{E}_n$ are pairs (H_k, ζ), of which ζ is an element of S, and $k = 1, 2, \ldots,$ or n. The universal set for E is the collection S' of all these pairs. Events in E are pairs (H_k, A), of which A is any event defined in S. The probability of the event (H_k, A) in E is

$$\Pr(H_k, A) = p_k \Pr(A \mid H_k) = p_k \Pr_k(A).$$

The event "hypothesis H_j is true" is (H_j, S), $j = 1, 2, \ldots, n$. Its probability p_j is the probability that experiment E_j is selected and is called the *prior probability*, or the *probability a priori*, of hypothesis H_j.

For the sake of definiteness let us suppose that any one of four mutually exclusive events A_1, A_2, A_3, and A_4 can occur, and that one of three hypotheses H_1, H_2, and H_3 may be true ($n = 3$). The hypotheses have prior probabilities p_1, p_2, and p_3; $p_1 + p_2 + p_3 = 1$. There are twelve possibilities, which we list in the following table:

$$\begin{array}{cccc}
(H_1, A_1) & (H_1, A_2) & (H_1, A_3) & (H_1, A_4) \\
(H_2, A_1) & (H_2, A_2) & (H_2, A_3) & (H_2, A_4) \\
(H_3, A_1) & (H_3, A_2) & (H_3, A_3) & (H_3, A_4)
\end{array}$$

In the next table we list the probabilities of these events in the composite experiment $\mathsf{E} = p_1\mathsf{E}_1 + p_2\mathsf{E}_2 + p_3\mathsf{E}_3$. These probabilities add to 1. Here $\Pr_i(A_j)$ is the probability of the event A_j in experiment E_i, $i = 1, 2, 3$; $j = 1, 2, 3, 4$.

$$\begin{array}{cccc}
p_1 \Pr_1(A_1) & p_1 \Pr_1(A_2) & p_1 \Pr_1(A_3) & p_1 \Pr_1(A_4) \\
p_2 \Pr_2(A_1) & p_2 \Pr_2(A_2) & p_2 \Pr_2(A_3) & p_2 \Pr_2(A_4) \\
p_3 \Pr_3(A_1) & p_3 \Pr_3(A_2) & p_3 \Pr_3(A_3) & p_3 \Pr_3(A_4)
\end{array}$$

The sum of the probabilities in the top row equals the prior probability p_1 of hypothesis H_1; the sum of those in the second row equals p_2; and so on. The sum

$$p_1 \Pr_1(A_1) + p_2 \Pr_2(A_1) + p_3 \Pr_3(A_1) = \Pr(A_1)$$

of the probabilities in the first column equals the total probability of event A_1 in the composite experiment, and likewise for the rest of the columns.

In a given trial of the experiment E our observer perceives only which of the four events A_1, A_2, A_3, or A_4 occurs. On the basis of that observation he must decide which hypothesis H_1, H_2, or H_3 is true, that is, which of the component experiments E_1, E_2, or E_3 was carried out, and he would like to make correct decisions with the greatest possible probability. His decision strategy can be indicated by putting a box around one of the three probabilities in each column, as we have done in the table below:

$$\begin{array}{cccc}
\boxed{p_1 \Pr_1(A_1)} & p_1 \Pr_1(A_2) & p_1 \Pr_1(A_3) & p_1 \Pr_1(A_4) \\
p_2 \Pr_2(A_1) & \boxed{p_2 \Pr_2(A_2)} & p_2 \Pr_2(A_3) & \boxed{p_2 \Pr_2(A_4)} \\
p_3 \Pr_3(A_1) & p_3 \Pr_3(A_2) & \boxed{p_3 \Pr_3(A_3)} & p_3 \Pr_3(A_4)
\end{array}$$

Thus when event A_1 is observed to have occurred, hypothesis H_1 is selected; when A_2 or A_4 is observed, hypothesis H_2; and when A_3 is observed, hypothesis H_3. One entry in each column must have a box. If when hypothesis H_1 is true, event A_1 occurs, the observer will correctly choose hypothesis H_1; if A_2, A_3, or A_4 occurs, other hypotheses will be selected, and an error will be made.

The probability Q of correct decision is therefore the sum of the probabilities in the boxes; the sum of the unboxed probabilities is the probability $P_e = 1 - Q$ of error. If the probability Q of correct decision is to be maximum, the box should be placed around the largest of the three entries in each column. Thus when event A_1 occurs, the observer should decide for hypothesis H_1 if

$$p_1 \Pr_1(A_1) > p_2 \Pr_2(A_1) \quad \text{and} \quad p_1 \Pr_1(A_1) > p_3 \Pr_3(A_1).$$

If two or more entries in a column are equal and larger than any of the others, it does not matter which among the corresponding hypotheses is chosen.

If we divide each element in the first column by the total probability

$$p_1 \Pr_1(A_1) + p_2 \Pr_2(A_1) + p_3 \Pr_3(A_1) = \Pr(A_1)$$

of event A_1, those elements become the conditional or *posterior* probabilities $\Pr(H_i \mid A_1)$ of the three hypotheses, $i = 1, 2, 3$, and we see that hypothesis H_1 should be selected if $\Pr(H_1 \mid A_1)$ exceeds both $\Pr(H_2 \mid A_1)$ and $\Pr(H_3 \mid A_1)$. The strategy that maximizes the probability Q of correct decision and minimizes the probability P_e of error is the strategy that chooses the hypothesis with the largest conditional probability, given the event observed.

However many hypotheses H_i and mutually exclusive observable events A_j there may be in the composite experiment

$$\mathsf{E} = p_1\mathsf{E}_1 + p_2\mathsf{E}_2 + \cdots + p_n\mathsf{E}_n,$$

we can construct a similar table and define our decision strategy by boxing one and only one entry in each column. We maximize the probability Q of correct decision by putting a box around the entry having the greatest probability in each column. The maximum attainable probability Q of correct decision will then be the sum of the probabilities in the boxes. Because the conditional or posterior probability of each hypothesis is obtained by dividing each entry by the sum of the probabilities in its own column,

$$\Pr(H_j \mid A_k) = \frac{p_j \Pr_j(A_k)}{\Pr(A_k)}, \qquad \Pr(A_k) = \sum_{i=1}^{n} p_i \Pr_i(A_k),$$

the best strategy picks that hypothesis having the greatest among these conditional probabilities. The directive to maximize the probability Q of correct decision in this way is known as *Bayes's rule* [4]. We illustrate it here with two examples, and in the next section we show how it can be applied to the design of a communication system.

EXAMPLE 1.22

Looking back at Example 1.8, let us denote by H_A the hypothesis that a transistor picked at random out of a bin came from supplier A, by H_B the hypothesis that it came from supplier B, and by H_C the hypothesis that it came from supplier C. As in Example 1.8, the prior probabilities of these hypotheses are

$$p_a = 0.45, \qquad p_b = 0.25, \qquad p_c = 0.3.$$

The transistor is installed in some equipment and checked after 1000 hours' operation. Let M denote the event that it is found to have failed and M' the event that it is still working. From the conditional probabilities in Example 1.8,

$$\Pr(M \mid A) = 0.15, \qquad \Pr(M \mid B) = 0.40, \qquad \Pr(M \mid C) = 0.25,$$

we can make the tables of events in the composite experiment and of their probabilities:

		M	M'
(H_A, M)	(H_A, M')	0.0675	$\boxed{0.3825}$
(H_B, M)	(H_B, M')	$\boxed{0.10}$	0.15
(H_C, M)	(H_C, M')	0.075	0.225

For example, $\Pr(H_A, M) = p_a \Pr(M \mid A) = 0.45 \cdot 0.15 = 0.0675$. The probabilities in the table add to 1. We have put a box around the largest probability in each column. Bayes's rule directs us to choose hypothesis H_A, "The transistor came from supplier A," if we find it working and to choose hypothesis H_B, "The transistor came from supplier B," if we find it failed. The probability of correct decision is the sum of the probabilities in the boxes,

$$Q = 0.10 + 0.3825 = 0.4825$$

and the probability of error is $P_e = 1 - Q = 0.5175$.

The total probability that the transistor will have failed after 1000 hours is the sum of the probabilities in the column headed "M",

$$\Pr(M) = 0.0675 + 0.10 + 0.075 = 0.2425$$

as before, and the probability that it will be working is

$$\Pr(M') = 1 - 0.2425 = 0.7575.$$

The conditional probability that the transistor came from supplier A, given it has failed by the time we check it, is

$$\Pr(H_A \mid M) = \frac{0.0675}{0.2425} = 0.27835,$$

as in Example 1.9, and so on for the rest of the events in the table.

EXAMPLE 1.23

A coin is provided that may be biased; the probability p that it comes up heads may be any of the four multiples of 0.2 from 0.2 through 0.8, all four possibilities having the same prior probability $\frac{1}{4}$. We can imagine the coin to have been drawn at random from a large heap of coins of four types, indistinguishable to the eye and present in equal proportions. When a coin of type k is tossed, it comes up heads with probability $\pi_k = 0.2k$, $k = 1, 2, 3, 4$. The selected coin is tossed ten times, and three times it shows heads, seven times tails. Which is the most likely of the four possible values of the probability p for this coin?

Now hypothesis H_k states that a coin of type k was drawn; its prior probability is $\frac{1}{4}$. If we make a table of the kind just described, it will have eleven columns and four rows. The column corresponding to our observation of three heads and seven tails will contain the probabilities

$$H_1\colon \frac{1}{4}\binom{10}{3}\pi_1^3(1-\pi_1)^7 = \quad 0.050332$$

$$H_2\colon \frac{1}{4}\binom{10}{3}\pi_2^3(1-\pi_2)^7 = \quad \boxed{0.053748}$$

$$H_3\colon \frac{1}{4}\binom{10}{3}\pi_3^3(1-\pi_3)^7 = \quad 0.010617$$

$$H_4\colon \frac{1}{4}\binom{10}{3}\pi_4^3(1-\pi_4)^7 = \quad 0.000197.$$

The total probability of getting three heads and seven tails in ten tosses of the coin equals the sum of these probabilities, Pr (3H, 7T) = 0.11489. The largest of these probabilities is the second, and Bayes's rule directs us to decide for hypothesis H_2, "The coin we tossed is of type 2," whenever we observe three heads and seven tails.

Because the prior probabilities p_k of the four hypotheses are all equal to $\frac{1}{4}$ and because the binomial coefficients are the same in all entries of a column, deciding merely requires finding the largest among the four quantities

$$\pi_k^h(1-\pi_k)^{10-h}, \qquad 1 \le k \le 4,$$

where h is the number of heads. By comparing these quantities for successive values of h, you can see that the best strategy decides for hypothesis H_1 when $0 \le h \le 2$, for H_2 when $3 \le h \le 4$, for H_3 when $6 \le h \le 7$, and for H_4 when $8 \le h \le 10$. When five heads appear, it does not matter whether one chooses H_2 or H_3; the probability Q of correct decision will be the same. Alternatively, when $h = 5$ one can toss an unbiased coin and choose H_2 if it comes up heads, H_3 if it comes up tails.

You should show that in the notation of Sec. 1-8 the maximum probability of correct decision attained by this strategy equals

$$\begin{aligned} Q &= p_1 \Pr(0 \le h \le 2 \mid H_1) + p_2 \Pr(3 \le h \le 4 \mid H_2) \\ &\quad + \tfrac{1}{2}[p_2 \Pr(h = 5 \mid H_2) + p_3 \Pr(h = 5 \mid H_3)] + p_3 \Pr(6 \le h \le 7 \mid H_3) \\ &\quad + p_4 \Pr(8 \le h \le 10 \mid H_4) \\ &= \tfrac{1}{4}[B(2, 10; 0.2) + B(5, 10; 0.4) - B(2, 10; 0.4) \\ &\quad + B(7, 10; 0.6) - B(5, 10; 0.6) + 1 - B(7, 10; 0.8)] \\ &= 0.62197. \end{aligned}$$

The probability of error equals $1 - Q = 0.37803$.

The techniques of statistical hypothesis testing are useful in the design of experiments in science and engineering, as when one must determine whether a certain heat treatment enhances the durability of machinery parts, or whether a fertilizer significantly raises the yield of a certain crop. Concepts of hypothesis testing underlie the design of efficient receivers of signals in communications and radar. Examples will be presented in the next section, and the theory of signal detection will be discussed more extensively in Sec. 7-5.

1-11 Conditional Probabilities in Digital Communications

A. Digital Communication System

A digital communication system consists of a source of digits, a transmitter, a channel, and a receiver, as shown in Fig. 1.12. The source produces a digit every T seconds, selected from a fixed alphabet of symbols, such as $\{0, 1\}$ or $\{A, B, C\}$. The symbols express messages of some sort. To each symbol of the alphabet is assigned a different signal, which the transmitter sends along the channel whenever that symbol is emitted by the source. The signals last no longer than T seconds. The receiver sees each signal attenuated, possibly distorted by the channel, and corrupted by random noise. We assume for simplicity that the signals, even if distorted, do not overlap when they reach the receiver. On the basis of what the receiver perceives, it makes a decision about which of the possible signals was transmitted, and it issues the associated digit. How it makes these decisions will be described later. The decisions may be wrong, and the stream of digits emerging from the receiver may contain errors.

The digits emitted by the source contain an element of unpredictability or randomness, by virtue of which they convey information. If the receiver knew exactly what digits were forthcoming, it would not need to observe the incoming signals, nor indeed would the transmitter need to send them. It is appropriate, therefore, to describe the output of the source in terms of probability. That output may consist of the letters and spaces in English text, or of numbers spewed out by a computer; or the English text or the computer output may have been transformed into a new sequence of symbols by some encoding process. The mechanism of the source will

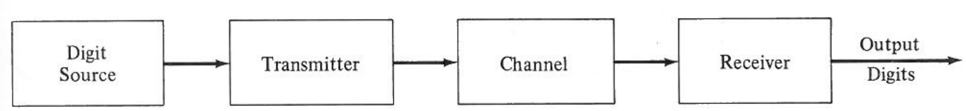

FIGURE 1.12 Digital communication system.

not concern us, and for simplicity we shall ignore relationships or "correlations" that may exist among the digits emerging from the source. Each instance of the process of digit emission, signal transmission, and receiver decision will be treated separately. We need therefore only to specify the probabilities with which the several digits are emitted by the source.

For definiteness let us consider binary messages in which the digits are 0's and 1's. Denote their probabilities by P_0 and P_1, $P_0 + P_1 = 1$. Roughly speaking, in a long stream of digits emitted by the source we should find a fraction P_0 of 0's and a fraction P_1 of 1's. They are to be transmitted by an optical communication system.

Whenever a 1 appears, the transmitter emits a pulse of light, which lasts no longer than T seconds and travels along an optical fiber to the receiver. Whenever the source produces a 0, no light pulse is sent. (This is sometimes called an "on–off" system.) At the far end of the fiber is a detector, which for simplicity we take to be a photoelectric cell. During the interval of T seconds' duration when the light pulse, if sent, would arrive, the receiver counts the number of electrons that have passed through the cell into the external circuitry. Denote this number by the integer k. It tends to be larger when a light pulse emerges from the fiber, the message digit having been a 1, than when no pulse has been transmitted, the message digit having been a 0. The electrons observed to pass through the photoelectric cell when no light is incident are known as "dark current." In either case the number k is not definite, but random; its value depends on whether electrons in the cathode, agitated by light and heat, happen to burrow to its surface, escape, and impinge on the anode. The random variability of the number k of electrons counted is a kind of noise. Again a probabilistic description is necessary.

The required probabilities must be specified in terms of the event structure of a certain chance experiment E. From the standpoint of an external observer it consists of the transmission or lack of transmission of a light pulse and the counting of so many electrons at the output of the photoelectric cell. It can be considered as the weighted sum

$$\mathsf{E} = P_0\mathsf{E}_0 + P_1\mathsf{E}_1$$

of two subexperiments E_0 and E_1. In subexperiment E_0 the number k of electrons emitted by the detector during T seconds is counted in the absence of any incident light. In subexperiment E_1 the number k of electrons is counted during the arrival of a T-second pulse of light from the transmitter. Which of these subexperiments is actually carried out in any interval is determined by the transmitter: E_0 when digit 0 is issued by the message source, E_1 when digit 1 is issued. The receiver does not know whether E_0 or E_1 was performed, but must decide on the basis of the number k of electrons it counts during the interval. It chooses between two hypotheses,

H_0: The message digit was 0; E_0 was performed.

H_1: The message digit was 1; E_1 was performed.

Subexperiment E_0 is selected by the transmitter with probability P_0, which is the prior probability of hypothesis H_0; subexperiment E_1 is selected with probability P_1, the prior probability of H_1.

The physical nature of the optical fiber and the photoelectric cell and the manner in which light emerging from the fiber impinges on the cell determine the probabilities $\Pr(k \mid H_0)$ and $\Pr(k \mid H_1)$ of counting k electrons in subexperiments E_0 and E_1, respectively. With respect to the composite experiment E these are conditional probabilities. In typical circumstances they will have the Poisson form

$$\Pr(k \mid H_1) = \frac{\lambda_1^k}{k!}e^{-\lambda_1}, \qquad \Pr(k \mid H_0) = \frac{\lambda_0^k}{k!}e^{-\lambda_0}, \tag{1.34}$$

with $\lambda_0 < \lambda_1$. They are sketched as bar graphs in Fig. 1.13. That the peak of the distribution $\Pr(k \mid H_1)$ lies to the right of that of $\Pr(k \mid H_0)$ corresponds to the statement that when a light pulse falls on the detector, more electrons are usually emitted than when one does not. By using the power series for e^x you can easily verify that, as for (1.27),

$$\sum_{k=0}^{\infty} \Pr(k \mid H_0) = \sum_{k=0}^{\infty} \Pr(k \mid H_1) = 1. \tag{1.35}$$

An outcome of the composite experiment E can be specified by a pair (H_i, k), where i is 1 or 0 depending on whether a pulse is or is not transmitted, and k is the number of electrons counted, $k = 0, 1, 2, \ldots$. Because these outcomes form a countable set, we can take each outcome (H_i, k) as an elementary event in E. By the definition of conditional probability the basic probability measure in the composite experiment E consists of probabilities of the form

$$\Pr(H_0, k) = P_0 \Pr(k \mid H_0) = \frac{P_0\lambda_0^k e^{-\lambda_0}}{k!} \tag{1.36}$$

$$\Pr(H_1, k) = P_1 \Pr(k \mid H_1) = \frac{P_1\lambda_1^k e^{-\lambda_1}}{k!} \tag{1.37}$$

for all k in $0 \le k < \infty$. All these probabilities add up to 1,

$$\sum_{i=0}^{1}\sum_{k=0}^{\infty} \Pr(H_i, k) = 1. \tag{1.38}$$

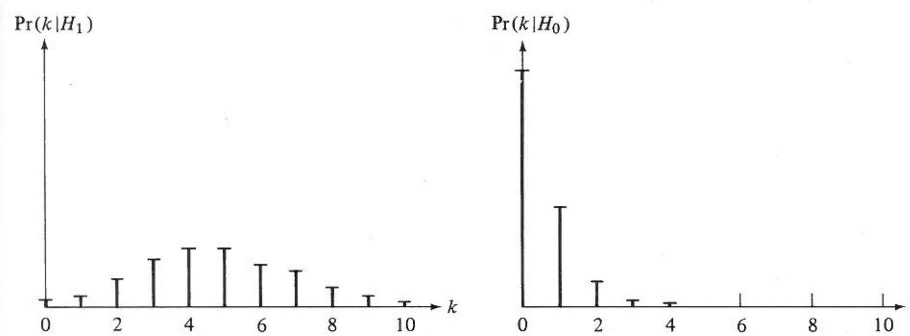

FIGURE 1.13 The conditional probabilities.

B. The Receiver as a Decision Mechanism

The receiver must decide every T seconds whether a pulse was transmitted or not, that is, whether the current message digit is a 0 or a 1. In the terminology of hypothesis testing, the receiver chooses between hypotheses H_0 and H_1. It must base its decision on the number k of electrons that it counts. As we did in the preceding section, we can make a table of the probabilities $\Pr(H_0, k)$ and $\Pr(H_1, k)$ for all k, putting the former into the top row, the latter into the bottom row. The kth column of that table will look like equations (1.36) and (1.37). We indicate our strategy by putting a box around the larger entry in each column in order to mark which hypothesis will be chosen. Thus if

$$P_1 \Pr(k \mid H_1) > P_0 \Pr(k \mid H_0), \tag{1.39}$$

we put the box around the bottom entry, and the receiver will select hypothesis H_1 and emit the digit 1 whenever it counts k photoelectrons. Otherwise we box the top entry, and the receiver emits a 0 for that value of the number k. This strategy, as we have seen, minimizes the probability P_e of error.

From (1.39) it follows that the receiver decides for hypothesis H_1 if

$$\frac{\Pr(k \mid H_1)}{\Pr(k \mid H_0)} > \frac{P_0}{P_1}. \tag{1.40}$$

The quantity on the left is called the *likelihood ratio*. From (1.34) it is given in this problem by

$$\frac{\Pr(k \mid H_1)}{\Pr(k \mid H_0)} = \left(\frac{\lambda_1}{\lambda_0}\right)^k \exp[-(\lambda_1 - \lambda_0)].$$

Hence the receiver decides for hypothesis H_1 and issues the digit 1 if

$$k > \frac{\lambda_1 - \lambda_0 + \ln(P_0/P_1)}{\ln(\lambda_1/\lambda_0)}; \tag{1.41}$$

otherwise it decides for H_0 and issues a 0.

We can represent this decision strategy as a division of the range $(0, \infty)$ of possible numbers k of counts into two parts, which we denote by R_0 and R_1. If k turns out to lie in R_0—we denote this by $k \in R_0$—the receiver decides that no pulse arrived and that the message digit is 0; if $k \in R_1$, it decides that a pulse arrived and that the message digit is 1. These "decision regions" are depicted in Fig. 1.14. The region R_0 consists of all integers k in $0 \le k \le \nu$, R_1 of integers k in $k > \nu$, where ν is an integral point of dichotomy given by (1.41) as

$$\nu = \left\lfloor \frac{\lambda_1 - \lambda_0 + \ln(P_0/P_1)}{\ln(\lambda_1/\lambda_0)} \right\rfloor, \tag{1.42}$$

where $\lfloor x \rfloor$ indicates the integral part of x.

In this communications receiver, a correct decision is made if a pulse was sent and $k \in R_1$ or if no pulse was sent and $k \in R_0$; this event consists of all outcomes

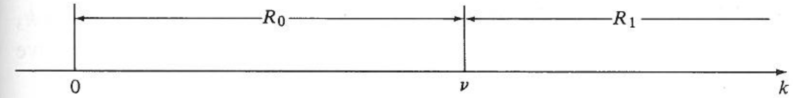

FIGURE 1.14 Decision regions.

(H_0, k) with $0 \le k \le \nu$ and all outcomes (H_1, k) with $k > \nu$. Its probability is

$$Q = \Pr(\text{correct decision}) = \sum_{k=0}^{\nu} \Pr(H_0, k) + \sum_{k=\nu+1}^{\infty} \Pr(H_1, k). \tag{1.43}$$

The complementary event is an error, and its probability is

$$P_e = \Pr(\text{error}) = \sum_{k=\nu+1}^{\infty} \Pr(H_0, k) + \sum_{k=0}^{\nu} \Pr(H_1, k). \tag{1.44}$$

Because we applied Bayes's rule in setting up our decision strategy, this receiver attains the largest possible probability Q of correct detection and the minimum probability P_e of error.

C. Ternary Communication System

The same principles apply to more complex communication systems. Suppose, for instance, that messages are encoded into an alphabet of three symbols $\{A, B, C\}$. The relative frequencies with which they occur are P_A, P_B, and P_C,

$$P_A + P_B + P_C = 1.$$

The transmitter has emitters of red, yellow, and blue light. Each A in the output of the digit source causes it to send a pulse of red light, each B a pulse of yellow light, and each C a pulse of blue light. At the far end of the optical fiber a prism imperfectly divides the light, passing most of the red light to detector 1, most of the yellow to detector 2, and most of the blue to detector 3. These three detectors are photoelectric cells, and during each pulse interval of duration T the receiver counts the number of electrons passing from each. Denote these three numbers by k_1, k_2, and k_3.

Now three subexperiments E_A, E_B, and E_C are at hand, and in the composite experiment

$$\mathsf{E} = P_A\mathsf{E}_A + P_B\mathsf{E}_B + P_C\mathsf{E}_C$$

the outcomes are denoted by (H_i, k_1, k_2, k_3), where $i = A, B,$ or C, and

$$0 \le k_1 < \infty, \qquad 0 \le k_2 < \infty, \qquad 0 \le k_3 < \infty.$$

The basic elements of the probability measure on E are the probabilities

$$\Pr(H_i, k_1, k_2, k_3) = P_i \Pr(k_1, k_2, k_3 \mid H_i), \qquad i = A, B, C;$$

$\Pr(k_1, k_2, k_3 \mid H_A)$ is the probability that detectors 1, 2, and 3 emit k_1, k_2, and k_3 photoelectrons, respectively, when a red pulse arrives along the fiber. It might have the form

$$\Pr(k_1, k_2, k_3 \mid H_A) = \frac{\lambda_{1A}^{k_1} \exp(-\lambda_{1A})}{k_1!} \frac{\lambda_{2A}^{k_2} \exp(-\lambda_{2A})}{k_2!} \frac{\lambda_{3A}^{k_3} \exp(-\lambda_{3A})}{k_3!} \tag{1.45}$$

for certain known constants $\lambda_{1A} > \lambda_{2A} > \lambda_{3A}$. The conditional probabilities $\Pr(k_1, k_2, k_3 \mid H_B)$ and $\Pr(k_1, k_2, k_3 \mid H_C)$ would be given by similar expressions.

The receiver must now choose among the three hypotheses labeled H_A, H_B, and H_C; H_A is the hypothesis "a red pulse was transmitted; subexperiment E_A is being performed," and so on. The receiver is equipped with a decision rule that divides the set of all triads (k_1, k_2, k_3) of nonnegative integers into three subsets R_A, R_B, and R_C. If $(k_1, k_2, k_3) \in R_A$, the receiver decides for hypothesis H_A and issues the digit A; if $(k_1, k_2, k_3) \in R_B$, it issues a B; if $(k_1, k_2, k_3) \in R_C$, a C. Again the events "correct decision" and "error" can be identified and their probabilities calculated:

$$\begin{aligned} \Pr(\text{correct decision}) = & \sum_{(k_1,k_2,k_3)\in R_A} \Pr(H_A, k_1, k_2, k_3) \\ & + \sum_{(k_1,k_2,k_3)\in R_B} \Pr(H_B, k_1, k_2, k_3) \\ & + \sum_{(k_1,k_2,k_3)\in R_C} \Pr(H_C, k_1, k_2, k_3) \end{aligned} \tag{1.46}$$

and $\Pr(\text{error}) = 1 - \Pr(\text{correct decision})$.

If the probability of error is to be minimum, or the probability of correct decision maximum, the receiver will apply Bayes's rule. That is, for the triad of numbers k_1, k_2, and k_3 of electrons counted at the three detectors, it will calculate the three posterior probabilities

$$\Pr(H_A \mid k_1, k_2, k_3), \quad \Pr(H_B \mid k_1, k_2, k_3), \quad \Pr(H_C \mid k_1, k_2, k_3).$$

It will decide for hypothesis H_A and issue the digit A if the first of these conditional probabilities is the largest of the three; if the second is largest, it issues digit B; and if the third is largest, digit C. This rule can be translated into a division of the set of all triads (k_1, k_2, k_3) into decision regions R_A, R_B, and R_C by using formulas like those in (1.45). The first posterior probability, for instance, is

$$\begin{aligned} \Pr(H_A \mid k_1, k_2, k_3) &= \frac{\Pr(H_A, k_1, k_2, k_3)}{\Pr(k_1, k_2, k_3)} \\ &= \frac{P_A \Pr(k_1, k_2, k_3 \mid H_A)}{\Pr(k_1, k_2, k_3)} \end{aligned} \tag{1.47}$$

with

$$\begin{aligned} \Pr(k_1, k_2, k_3) = & P_A \Pr(k_1, k_2, k_3 \mid H_A) \\ & + P_B \Pr(k_1, k_2, k_3 \mid H_B) + P_C \Pr(k_1, k_2, k_3 \mid H_C) \end{aligned}$$

by the principle of total probability. The remaining posterior probabilities are given by similar expressions. The region R_A consists of all triads for which

$$\Pr(H_A \mid k_1, k_2, k_3) > \Pr(H_B \mid k_1, k_2, k_3)$$

and

$$\Pr(H_A \mid k_1, k_2, k_3) > \Pr(H_C \mid k_1, k_2, k_3),$$

and R_B and R_C are similarly specified.

Alternatively, we can make an imaginary table like the one we constructed in Sec. 1-10. Each column will correspond to a triad (k_1, k_2, k_3) of the numbers of electrons counted in the three detectors. The entries in that column will be the three probabilities

$$P_i \Pr(k_1, k_2, k_3 \mid H_i), \qquad i = A, B, C.$$

The best strategy marks the largest of these three probabilities to indicate which hypothesis is chosen when that particular triad of counts occurs. Thus the receiver chooses hypothesis H_A and emits the symbol A if both

$$P_A \Pr(k_1, k_2, k_3 \mid H_A) > P_B \Pr(k_1, k_2, k_3 \mid H_B)$$

and

$$P_A \Pr(k_1, k_2, k_3 \mid H_A) > P_C \Pr(k_1, k_2, k_3 \mid H_C),$$

and so on for the other two possible orders of the numbers in the column (k_1, k_2, k_3) of the table. By this procedure we can sort all the triads among the decision regions R_A, R_B, and R_C. The probability of correct decision attained by this system will be given by (1.46) and will be as large as circumstances permit.

In setting up a receiver to make decisions by Bayes's rule, probability is being applied in the hypothetical mode defined at the beginning of Sec. 1-4. Just what sequence of digits will be issued by the message source, and how many photoelectrons each received pulse will induce in the detector, cannot be predicted. Any proposed strategy by which the receiver might make its decisions must be tested with respect to the entire ensemble of possible combinations of digits and photoelectron counts. In effect the number of errors that result is determined for each such combination and weighted in accordance with the probability that that combination occurs. The probabilities are drawn from a model of the communication system that takes into account the nature of message source, channel, and detector. Prior probabilities such as P_0 and P_1, or P_A, P_B, and P_C, may be estimated from a long sequence of message digits observed during some past interval, or from studies of the language expressing the messages and the manner of their encoding. Conditional probabilities such as $\Pr(k \mid 0)$ and $\Pr(k \mid 1)$ for the numbers k of photoelectrons counted are derived from the physics of channel and detector. The model is justified by the past experience that validated the scientific theories on which it is founded. The probability Pr (error) calculated from this model assesses the effectiveness of a given decision strategy, for among all combinations of very long sequences of N transmitted digits and the ensuing numbers of counts, by far the greatest probability will attach to those in which the number of errors is close to N Pr (error)—in a sense whose precise formulation must await further development of this subject. The designer builds into the receiver that strategy for which the probability Pr (error) is minimum.

PROBLEMS

1-1 For the events A, B, and C defined for experiment E_f on page 5, determine the events A', B', C', $A \cap B$, $A \cap C$, $B \cap C$, $A' \cap B$, $B' \cap C'$, $A \cup B$, $B \cup C$, and $B' \cup C'$. Be careful to distinguish "$<$" from "$\le$", "$>$" from "$\ge$."

1-2 Illustrate the rules (1.1) through (1.8) by means of Venn diagrams. Use colored pencils if helpful.

1-3 Referring to experiment E_g, represent on a diagram the events A, B, and C defined on page 5, and indicate on similar diagrams the events C', $A' \cap B$, $A \cap C$, $(A \cap B) \cup C$.

1-4 Let S be the set of positive integers $n = 1, 2, 3, \ldots$. Given the following three subsets of S,

$$\begin{aligned} A &= \{n : n \text{ is divisible by } 3\} = \{3, 6, 9, 12, \ldots\}, \\ B &= \{n : n \text{ is divisible by } 4\} = \{4, 8, 12, 16, \ldots\}, \\ C &= \{n : n \text{ is divisible by } 5\} = \{5, 10, 15, 20, \ldots\}, \end{aligned}$$

find the sets A', B', C', $A \cap B$, $A \cup B$, $A \cap C$, $A \cup C$, and $A \cap B \cap C$.

1-5 Let S be the upper-right quadrant of the XY-plane, that is,

$$S = \{(x, y) : x \ge 0,\ y \ge 0\}.$$

Given the following three subsets of S,

$$\begin{aligned} A &= \{(x, y) : x \ge 2y\}, \\ B &= \{(x, y) : x + y < 5\}, \\ C &= \{(x, y) : xy > 1\}, \end{aligned}$$

sketch in the XY-plane the sets A', B', C', $A \cap B$, $A \cup B$, $A \cap C$, $A \cup C$, and $A \cap B \cap C$.

1-6 Consider events in experiment E_g, page 4, of the form

$$\begin{aligned} A_0 &= \{(v_1, v_2) : v_1^2 + v_2^2 \le 1\}, \\ A_1 &= \{(v_1, v_2) : v_1^2 + v_2^2 \le \tfrac{1}{2}\}, \\ &\cdot \\ &\cdot \\ &\cdot \\ A_k &= \{(v_1, v_2) : v_1^2 + v_2^2 \le 2^{-k}\}, \ldots \end{aligned}$$

for all nonnegative integers k. Determine the events

$$\bigcap_{k=1}^{\infty} A_k \qquad \text{and} \qquad \bigcup_{k=1}^{\infty} A_k'.$$

1-7 Write a detailed proof of Corollary 4 by using Axioms I–III and the rules of set theory.

1-8 Derive the counterpart of Corollary 4 for three overlapping events A, B, and C, that is, for three events that may have some outcomes in common.

1-9 Show that for any n events

$$\Pr(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n) = \Pr\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \Pr(A_i).$$

Communication theorists call this the "union bound" and find it useful when the event on the left is too complicated for its probability to be calculated exactly.

1-10 Suppose that the intervals in Example 1.4 overlap: $a < c < b < d$. Determine Pr $(v : a \le v < b \text{ or } c \le v < d)$. Verify Corollary 4 for this example, taking $A = \{v : a \le v < b\}$, $B = \{v : c \le v < d\}$ and determining Pr (A), Pr (B), and Pr $(A \cap B)$ as well.

1-11 A child's game consists of spinning two arrows, each pivoted about its center of gravity. An arrow is just as likely to stop at one azimuth as at another, and where the one stops is independent of where the other stops. The board under the first arrow is divided into angular sectors numbered 1 through 4 and encompassing the following angles:

1: $0° < \theta < 80°$, 2: $80° < \theta < 150°$, 3: $150° < \theta < 270°$,
4: $270° < \theta < 360°$.

The board under the second arrow is similarly divided:

1: $0° < \theta < 120°$, 2: $120° < \theta < 200°$, 3: $200° < \theta < 310°$,
4: $310° < \theta < 360°$.

One's score is the sum of the numbers of the sectors in which the two arrows stop. Calculate the probability Pr (k) that one achieves a score of k points for all integers k from 2 through 8. Be sure that your answers add up to 1.

1-12 Under each of the assumptions (a), (b), and (c) in Sec. 1-4A, calculate the probability Pr $(0 \le \theta < \phi)$ that the angle θ subtended by the chord AB in Fig. 1.2 is less than ϕ, $0 \le \phi < \pi$, and sketch these probabilities as functions of the angle ϕ. (Bertrand's paradox is usually discussed for $\phi = 2\pi/3$, whereupon the chord AB is required to be shorter than the side of an equilateral triangle inscribed in the circle, but geometrical elegance is the only virtue of that special case.)

1-13 The probability that n electrons are emitted during a certain interval from a photoelectrically emissive surface under incident light is

$$\Pr(n) = Cv^n, \qquad 0 \le n < \infty, \qquad 0 < v < 1.$$

Calculate the constant C.