6

Infinite Sequences of Random Variables

6.1 INTRODUCTION

We have not yet encountered any situation in which it is necessary to consider an infinite collection of random variables, all defined on the same probability space. In the central limit theorem, for example, the basic underlying hypothesis is "For each n, let $R_1, \ldots, R_n$ be independent random variables." As n changes, the underlying probability space may change, but this is of no consequence, since a convergence in distribution statement is a statement about convergence of a sequence of real-valued functions on E^1. If $R_1, \ldots, R_n$ are independent, with distribution functions $F_1, \ldots, F_n$, and $T_n = (S_n - E(S_n))/\sigma(S_n)$, $S_n = R_1 + \cdots + R_n$, the distribution function of T_n is completely determined by the F_i, and the validity of a statement about convergence in distribution of T_n is also determined by the F_i, regardless of the construction of the underlying space.

However, there are occasions when it is necessary to consider an infinite number of random variables defined on the same probability space. For example, consider the following random experiment. We start at the origin on the real line, and flip a coin independently over and over again. If the result of the first toss is heads, we take one step to the right (i.e., from $x = 0$

to $x = 1$), and if the result is tails, we move one step to the left (to $x = -1$). We continue the process; if we are at $x = k$ after n trials, then at trial $n + 1$ we move to $x = k + 1$ if the $(n + 1)$th toss results in heads, or to $x = k - 1$ if it results in tails. We ask, for example, for the probability of eventually returning to the origin.

Now the position S_n after n steps is the sum $R_1 + \cdots + R_n$ of n independent random variables, where $P\{R_k = 1\} = p =$ probability of heads, $P\{R_k = -1\} = 1 - p$. We are looking for $P\{S_n = 0 \text{ for some } n > 0\}$.

We must decide what probability space we are considering. If we are interested only in the first n trials, there is no problem. We simply have a sequence of n Bernoulli trials, and we have considered the assignment of probabilities in detail. However, the difficulty is that the event $\{S_n = 0$ for some $n > 0\}$ involves infinitely many trials. We must take $\Omega = E^\infty =$ all *infinite* sequences $(x_1, x_2, \ldots)$ of real numbers. (In this case we may restrict the x_i to be ± 1, but it is convenient to allow arbitrary x_i so that the discussion will apply to the general problem of assigning probabilities to events involving infinitely many random variables.)

We have the problem of specifying the sigma field $\mathscr{F}$ and the probability measure P. The physical description of the problem has determined all probabilities involving finitely many R_i; that is, we know $P\{(R_1, \ldots, R_n) \in B\}$ for each positive integer n and n-dimensional Borel set B. What we would like to conclude is that a reasonable specification of probabilities involving finitely many R_i determines the probability of events involving all the R_i. For example, consider $\{\text{all } R_i = 1\}$. This event may be expressed as

$$\bigcap_{n=1}^{\infty} \{R_1 = 1, \ldots, R_n = 1\}$$

The sets $\{R_1 = 1, \ldots, R_n = 1\}$ form a contracting sequence; hence

$$P\{\text{all } R_i = 1\} = \lim_{n\to\infty} P\{R_1 = 1, \ldots, R_n = 1\} = \lim_{n\to\infty} p^n = 0 \qquad \text{if } p < 1$$

As another example,

$$\begin{aligned}\{R_n = 1 \text{ for infinitely many } n\} \\ &= \{\text{for every } n, \text{ there exists } k \geq n \text{ such that } R_k = 1\} \\ &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{R_k = 1\}\end{aligned}$$

Thus

$$\begin{aligned}P\{R_n = 1 \text{ for infinitely many } n\} &= \lim_{n\to\infty} P\left[\bigcup_{k=n}^{\infty} \{R_k = 1\}\right] \\ &= \lim_{n\to\infty} \lim_{m\to\infty} P\left[\bigcup_{k=n}^{m} \{R_k = 1\}\right]\end{aligned}$$

Thus again the probability is determined once we know the probabilities of all events involving finitely many R_i.

We now sketch the general situation. Let $\Omega = E^\infty$. A set of the form $\{(x_1, x_2, \ldots): (x_1, \ldots, x_n) \in B_n\}$, where $B_n \subset E^n$, is called a *cylinder* with *base* B_n, a *measurable cylinder* if B_n is a Borel subset of E^n.

Suppose that for each n we specify a probability measure P_n on the Borel subsets of E^n; $P_n(B_n)$ is to be interpreted as $P\{(R_1, \ldots, R_n) \in B_n\}$, where $R_i(x_1, x_2, \ldots) = x_i$.

Suppose, for example, that we have specified P_5. Then P_k, $k < 5$, is determined. In particular, in the discrete case we have

$$P\{(R_1, R_2, R_3) \in B_3\} = \sum_{\substack{(x_1,x_2,x_3)\in B_3, \\ -\infty < x_4 < \infty, \\ -\infty < x_5 < \infty}} P\{R_1 = x_1, R_2 = x_2, R_3 = x_3, R_4 = x_4, R_5 = x_5\}$$

and in the absolutely continuous case we have

$$P\{(R_1, R_2, R_3) \in B_3\} = \int \cdots \int_{\substack{(x_1,x_2,x_3)\in B_3 \\ -\infty < x_4 < \infty, \\ -\infty < x_5 < \infty}} f(x_1, x_2, x_3, x_4, x_5)\, dx_1\, dx_2\, dx_3\, dx_4\, dx_5$$

In general, once P_n is given, P_k, $k < n$, is determined. But we have specified P_k, $k < n$, at the beginning; if our assignment of probabilities is to make sense, the original P_k must agree with that derived from P_n, $n > k$.

If, for all $n = 1, 2, \ldots$ and all $k < n$, the probability measure P_k originally specified agrees with that derived from P_n, $n > k$, we say that the probability measures are *consistent*. Under the consistency hypothesis, the *Kolmogorov extension theorem* states that there is a unique probability measure P on $\mathscr{F}$ = the smallest sigma field of subsets of Ω containing the measurable cylinders, such that

$$P\text{ (the measurable cylinder with base } B_n) = P_n(B_n)$$

for all $n = 1, 2, \ldots$ and all Borel subsets B_n of E^n.

In other words, a consistent specification of finite dimensional probabilities determines the probabilities of events involving all the R_i.

We now consider the case in which the R_i are discrete. Here we determine probabilities involving $(R_1, \ldots, R_n)$ by prescribing the joint probability function

$$p_{12\cdots n}(x_1, \ldots, x_n) = P\{R_1 = x_1, \ldots, R_n = x_n\}$$

We may then derive the joint probability function of $R_1, \ldots, R_k$:

$$P\{R_1 = x_1, \ldots, R_k = x_k\} = \sum_{x_{k+1}, \ldots, x_n} P\{R_1 = x_1, \ldots, R_n = x_n\} \quad (6.1.1)$$

If this coincides with the given $p_{12\cdots k}$ (for all n and all $k < n$) we say that the system of joint probability functions is consistent. If we sum (6.1.1) over $(x_1, \ldots, x_k) \in B_k$, we find that consistency of the joint probability functions is equivalent to consistency of the associated probability measures P_n. Thus in the discrete case the essential point is the consistency of the joint probability functions. In particular, suppose that we require that for each n, $R_1, \ldots, R_n$ be independent, with R_i having a specified probability function p_i. Then (6.1.1) becomes

$$P\{R_1 = x_1, \ldots, R_k = x_k\} = \sum_{x_{k+1}, \ldots, x_n} p_1(x_1) \cdots p_n(x_n) = p_1(x_1) \cdots p_k(x_k)$$

and thus the joint probability functions are consistent. The point we are making here is that there is a unique probability measure on $\mathscr{F}$ such that the random variables $R_1, R_2, \ldots$ are independent, each with a specified probability function. In other words, the statement "Let $R_1, R_2, \ldots$ be independent random variables, where R_i is discrete and has probability function p_i," is unambiguous.

In the absolutely continuous case, probabilities involving $R_1, \ldots, R_n$ are determined by the joint density function $f_{12\cdots n}$. The joint density of $R_1, \ldots, R_k$ is then given by

$$g(x_1, \ldots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{12\cdots n}(x_1, \ldots, x_n)\, dx_{k+1} \cdots dx_n \tag{6.1.2}$$

If this coincides with the given $f_{12\cdots k}$ $(n, k = 1, 2, \ldots, k < n)$, we say that the system of joint densities is consistent. By integrating (6.1.2) over Borel sets $B_k \subset E^k$, we find that consistency of joint density functions is equivalent to consistency of the associated probability measures P_n. In particular, if we require that for each n, $R_1, \ldots, R_n$ be independent, with R_i having a specified density function f_i, then (6.1.2) becomes

$$\begin{aligned} g(x_1, \ldots, x_k) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_1(x_1) \cdots f_n(x_n)\, dx_{k+1} \cdots dx_n \\ &= f_1(x_1) \cdots f_k(x_k) \end{aligned}$$

Therefore the joint density functions are consistent, and the statement "Let $R_1, R_2, \ldots$ be independent random variables, where R_i is absolutely continuous with density function f_i," is unambiguous.

PROBLEMS

1. By working directly with the probability measures P_n, give an argument shorter than the one above to show that the statement "Let $R_1, R_2, \ldots$ be independent random variables, where R_i is absolutely continuous with density f_i," is unambiguous.

2. If $R_1, R_2, \dots$ are independent, with $P\{R_i = 1\} = p$, $P\{R_i = -1\} = 1 - p$, as at the beginning of the section, find $P\{R_n = 1 \text{ for infinitely many } n\}$; also, find $P\{\lim_{n\to\infty} R_n = 1\}$. (Assume $0 < p < 1$.)

6.2 THE GAMBLER'S RUIN PROBLEM

Suppose that a gambler starts with a capital of x dollars and plays a sequence of games against an opponent with $b - x$ dollars. At each trial he wins a dollar with probability p, and loses a dollar with probability $q = 1 - p$. (The trials are assumed independent, with $0 < p < 1$, $0 < x < b$.) The process continues until the gambler's capital reaches 0 (ruin) or b (victory). We wish to find $h(x)$, the probability of eventual ruin when the initial capital is x.

Let $A = \{\text{eventual ruin}\}$, $B_1 = \{\text{win on trial 1}\}$, $B_2 = \{\text{lose on trial 1}\}$. By the theorem of total probability,

$$P(A) = P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2)$$

We are given that $P(B_1) = p$, $P(B_2) = q$; $P(A)$ is the unknown probability $h(x)$. Now if the gambler wins at the first trial, his capital is then $x + 1$; thus $P(A \mid B_1)$ is the probability of eventual ruin, starting at $x + 1$, that is, $h(x + 1)$. Similarly, $P(A \mid B_2) = h(x - 1)$. Thus

$$h(x) = ph(x + 1) + qh(x - 1),\ x = 1, 2, \dots, b - 1 \tag{6.2.1}$$

[The intuition behind the argument leading to (6.2.1) is compelling; however, a formal proof involves concepts not treated in this book, and will be omitted.]

We have not yet found $h(x)$, but we know that it satisfies (6.2.1), a linear homogeneous *difference equation* with constant coefficients. The boundary conditions are $h(0) = 1$, $h(b) = 0$. To see this, note that if $x = 1$, then with probability p the gambler wins on trial 1; his probability of eventual ruin is then $h(2)$. With probability q he loses on trial 1, and then he is already ruined. In other words, if (6.2.1) is to be satisfied at $x = 1$, we must have $h(0) = 1$. Similarly, examination of (6.2.1) at $x = b - 1$ shows that $h(b) = 0$.

The difference equation may be put into the standard form

$$ph(x + 2) - h(x + 1) + qh(x) = 0,$$
$$x = 0, 1, \dots, b - 2,\ h(0) = 1,\ h(b) = 0$$

It is solved in the same way as the analogous differential equation

$$p\frac{d^2y}{dx^2} - \frac{dy}{dx} + qy = 0$$

We assume an exponential solution; for convenience, we take $h(x) = \lambda^x$ $(= e^{x \ln \lambda})$. Then $p\lambda^{x+2} - \lambda^{x+1} + q\lambda^x = \lambda^x(p\lambda^2 - \lambda + q) = 0$. Since λ^x is never 0, the only allowable λ's are the roots of the *characteristic equation* $p\lambda^2 - \lambda + q = 0$, namely,

$$\lambda = \frac{1}{2p}(1 \pm \sqrt{1 - 4pq})$$

Now

$$\begin{aligned}(p - q)^2 = p^2 - 2pq + q^2 &= p^2 + 2pq + q^2 - 4pq = (p + q)^2 - 4pq \\ &= 1 - 4pq \end{aligned} \tag{6.2.2}$$

Hence

$$\lambda = \frac{1}{2p}(1 \pm |p - q|)$$

The two roots are

$$\lambda_1 = \frac{1 + p - q}{2p} = 1, \qquad \lambda_2 = \frac{1 + q - p}{2p} = \frac{q}{p}$$

CASE 1. $p \neq q$. Then λ_1 and λ_2 are distinct; hence

$$h(x) = A\lambda_1^x + C\lambda_2^x = A + C\left(\frac{q}{p}\right)^x$$

$$h(0) = A + C = 1$$

$$h(b) = A + C\left(\frac{q}{p}\right)^b = 0$$

Solving, we obtain

$$A = \frac{-(q/p)^b}{1 - (q/p)^b} \qquad C = \frac{1}{1 - (q/p)^b}$$

Therefore

$$h(x) = \frac{(q/p)^x - (q/p)^b}{1 - (q/p)^b} \tag{6.2.3}$$

CASE 2. $p = q = 1/2$. Then $\lambda_1 = \lambda_2 = \lambda = 1$, a repeated root. In such a case (just as in the analogous differential equation) we may construct two linearly independent solutions by taking λ^x and $x\lambda^x$; that is,

$$h(x) = A\lambda^x + Cx\lambda^x = A + Cx$$

$$h(0) = A = 1$$

$$h(b) = A + Cb = 0 \qquad \text{so } C = -\frac{1}{b}$$

Thus

$$h(x) = 1 - \frac{x}{b} = \frac{b - x}{b} \tag{6.2.4}$$

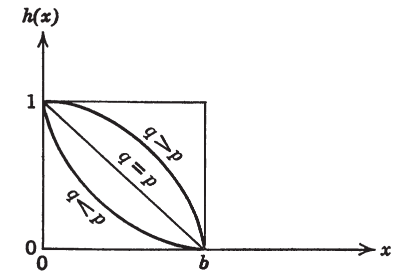

FIGURE 6.2.1 Probability of Eventual Ruin.

so that the probability of eventual ruin is the ratio of the adversary's capital to the total capital. A sketch of $h(x)$ in the various cases is shown in Figure 6.2.1.

Similarly, let $g(x)$ be the probability of eventual victory, starting with a capital of x dollars. We cannot conclude immediately that $g(x) = 1 - h(x)$, since there is the possibility that the game will never end; that is, the gambler's fortune might oscillate forever within the limits $x = 1$ and $x = b - 1$. However, we can show that this event has probability 0, as follows. By the same reasoning as that leading to (6.2.1), we obtain

$$g(x) = pg(x + 1) + qg(x - 1) \tag{6.2.5}$$

The boundary conditions are now $g(0) = 0$, $g(b) = 1$. But we may verify that $g(x) = 1 - h(x)$ satisfies (6.2.5) with the given boundary conditions; since the solution is unique (see Problem 1), we must have $g(x) = 1 - h(x)$; that is, *the game ends with probability 1.*

We should mention, at least in passing, the probability space we are working in. We take $\Omega = E^\infty$, $\mathscr{F} =$ the smallest sigma field containing the measurable cylinders, $R_i(x_1, x_2, \ldots) = x_i$, $i = 1, 2, \ldots$, P the probability measure determined by the requirement that $R_1, R_2, \ldots$ be independent, with $P\{R_i = 1\} = p$, $P\{R_i = -1\} = q$. Thus R_i is the gambler's net gain on trial i, and $x + \sum_{i=1}^n R_i$ is his capital after n trials. We are looking for $h(x) = P\{\text{for some } n,\ x + \sum_{i=1}^n R_i = 0,\ 0 < x + \sum_{i=1}^k R_i < b,\ k = 1, 2, \ldots, n - 1\}$.

A sequence of random variables of the form $x + \sum_{i=1}^n R_i$, $n = 1, 2, \ldots$, where the R_i are independent and have the same distribution function (or, more generally, $R_0 + \sum_{i=1}^n R_i$, $n = 1, 2, \ldots$, where $R_0, R_1, R_2, \ldots$ are independent and $R_1, R_2, \ldots$ have the same distribution function), is called a *random walk*, a *simple* random walk if $R_i (i \geq 1)$ takes on only the values ± 1. The present case may be regarded as a simple random walk with *absorbing* *barriers* at 0 and b, since when the gambler's fortune reaches either of these figures, the game ends, and we may as well regard his capital as forever frozen.

We wish to investigate the effect of removing one or both of the barriers. Let $h_b(x)$ be the probability of eventual ruin starting from x, when the total capital is b. It is reasonable to expect that $\lim_{b\to\infty} h_b(x)$ should be the probability of eventual ruin when the gambler has the misfortune of playing against an adversary with infinite capital. Let us verify this.

Consider the simple random walk with only the barrier at $x = 0$ present; that is, the adversary has infinite capital. If the gambler starts at $x > 0$, his probability $h^*(x)$ of eventual ruin is

$$P(A) = P\{\text{for some positive integer } b, 0 \text{ is reached before } b\}$$

Let $A_b = \{0 \text{ is reached before } b\}$. The sets A_b, $b = 1, 2, \ldots$ form an expanding sequence whose union is A; hence

$$P(A) = \lim_{b\to\infty} P(A_b)$$

But

$$P(A_b) = h_b(x)$$

Consequently

$$\begin{aligned} h^*(x) &= \lim_{b\to\infty} h_b(x) \\ &= 1 \qquad \text{if } q \geq p \\ &= \left(\frac{q}{p}\right)^x \qquad \text{if } q < p, \text{ by (6.2.3) and (6.2.4) } (x = 1, 2, \ldots) \end{aligned} \tag{6.2.6}$$

Thus, in fact, $\lim_{b\to\infty} h_b(x)$ is the probability $h^*(x)$ of eventual ruin when the adversary has infinite capital; $1 - h^*(x)$ is the probability that the origin will never be reached, that is, that the game will never end. If $q < p$, then $h^*(x) < 1$, and so there is a positive probability that the game will go on forever.

Finally, consider a simple random walk starting at 0, with no barriers. Let r be the probability of eventually returning to 0. Now if $R_1 = 1$ (a win on trial 1), there will be a return to 0 with probability $h^*(1)$. If $R_1 = -1$ (a loss on trial 1), the probability of eventually reaching 0 is found by evaluating $h^*(1)$ with q and p interchanged, that is, 1 for $q \leq p$, and p/q if $p < q$.

Thus, if $q \leq p$,

$$r = p\left(\frac{q}{p}\right) + q(1) = 2q$$

If $p < q$,

$$r = p(1) + q\left(\frac{p}{q}\right) = 2p$$

One expression covers both of these cases, namely,

$$\begin{aligned} r &= 1 - |p - q| \\ &= 1 \qquad \text{if } p = q = \tfrac{1}{2} \\ &< 1 \qquad \text{if } p \neq q \end{aligned} \tag{6.2.7}$$

PROBLEMS

1. Show that the difference equation arising from the gambler's ruin problem has a unique solution subject to given boundary conditions at $x = 0$ and $x = b$.

2. In the gambler's ruin problem, let $D(x)$ be the average duration of the game when the initial capital is x. Show that $D(x) = p(1 + D(x + 1)) + q(1 + D(x - 1))$, $x = 1, 2, \ldots, b - 1$ [the boundary conditions are $D(0) = D(b) = 0$].

3. Show that the solution to the difference equation of Problem 2 is

$$\begin{aligned} D(x) &= \frac{x}{q - p} - \frac{(b/(q - p))(1 - (q/p)^x)}{1 - (q/p)^b} \qquad \text{if } p \neq q \\ &= x(b - x) \qquad \text{if } p = q = 1/2 \end{aligned}$$

[$D(x)$ can be shown to be finite, so that the usual method of solution applies; see Problem 4, Section 7.4.]

REMARK. If $D_b(x)$ is the average duration of the game when the total capital is b, then $\lim_{b\to\infty} D_b(x)$ ($= \infty$ if $p \geq q$, $= x/q - p$ if $p < q$) can be interpreted as the average length of time required to reach 0 when the adversary has infinite capital.

4. In a simple random walk starting at 0 (with no barriers), show that the average length of time required to return to the origin is infinite. (Corollary: A couple decides to have children until the number of boys equals the number of girls. The average number of children is infinite.)

5. Consider the simple random walk starting at 0. If $b > 0$, find the probability that $x = b$ will eventually be reached.

6.3 COMBINATORIAL APPROACH TO THE RANDOM WALK; THE REFLECTION PRINCIPLE

In this section we obtain, by combinatorial methods, some explicit results connected with the simple random walk. We assume that the walk starts at 0, with no barriers; thus the position at time n is $S_n = \sum_{k=1}^{n} R_k$, where $R_1, R_2, \ldots$ are independent random variables with $P\{R_k = 1\} = p$,

$P\{R_k = -1\} = q = 1 - p$. We may regard the R_k as the results of an infinite sequence of Bernoulli trials; we call an occurrence of $R_k = 1$ a "success," and that of $R_k = -1$ a "failure."

Suppose that among the first n trials there are exactly a successes and b failures $(a + b = n)$; say $a > b$. We ask for the (conditional) probability that the process will always be positive at times $1, 2, \ldots, n$, that is,

$$P\{S_1 > 0, S_2 > 0, \ldots, S_n > 0 \mid S_n = a - b\} \tag{6.3.1}$$

(Notice that the only way that S_n can equal $a - b$ is for there to be a successes and b failures in the first n trials; for if x is the number of successes and y the number of failures, then $x + y = n = a + b$, $x - y = a - b$; hence $x = a$, $y = b$. Thus $\{S_n = a - b\} = \{a$ successes, b failures in the first n trials$\}$.)

Now (6.3.1) may be written as

$$\frac{P\{S_1 > 0, \ldots, S_n > 0, S_n = a - b\}}{P\{S_n = a - b\}} \tag{6.3.2}$$

A favorable outcome in the numerator corresponds to a path from (0, 0) to $(n, a - b)$ that always lies above the axis,† and a favorable outcome in the denominator to an arbitrary path from (0, 0) to $(n, a - b)$ (see Figure 6.3.1). Thus (6.3.2) becomes

$$\frac{p^a q^b[\text{the number of paths from } (0,0) \text{ to } (n, a-b) \text{ that are always above } 0]}{p^a q^b[\text{the total number of paths from } (0,0) \text{ to } (n, a-b)]}$$

A path from (0, 0) to $(n, a - b)$ is determined by selecting a positions out of n for the successes to occur; the total number of paths is $\binom{n}{a} = \binom{a+b}{a}$. To count the number of paths lying above 0, we reason as follows (see Figure 6.3.2).

Let A and B be points above the axis. Given any path from A to B that touches or crosses the axis, reflect the segment between A and the first zero point T, as shown. We get a path from A' to B, where A' is the reflection of A. Conversely, given *any* path from A' to B, the path must reach the axis at some point T. Reflecting the segment from A' to T, we obtain a path from A to B that touches or crosses the axis. The correspondence thus established is one-to-one; hence

the number of paths from A to B that touch or cross the axis

$=$ *the total number of paths from A' to B*

† Terminology: For the purpose of determining whether or not a path lies above the axis (or touches it, crosses it, etc.), the end points are not included in the path.

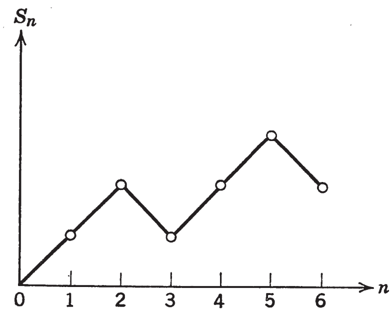

FIGURE 6.3.1 A Path in the Random Walk.

$$n = 6$$

$$a = 4, \qquad b = 2$$

$$P\{R_1 = R_2 = 1,\ R_3 = -1,\ R_4 = R_5 = 1,\ R_6 = -1\} = p^4q^2;$$

this is one contribution to

$$P\{S_1 > 0, \ldots, S_6 > 0, S_6 = 2\}$$

This is called the *reflection principle*. Now

the number of paths from $(0, 0)$ to $(n, a - b)$ lying entirely above the axis

= the number from $(1, 1)$ to $(n, a - b)$ that *neither touch nor cross* the axis (since R_1 must be $+1$ in this case)

= the total number from $(1, 1)$ to $(n, a - b)$ − the number from $(1, 1)$ to $(n, a - b)$ that either touch or cross the axis

= the total number from $(1, 1)$ to $(n, a - b)$ − the total number from $(1, -1)$ to $(n, a - b)$ (by the reflection principle)

$$= \binom{n-1}{a-1} - \binom{n-1}{a}$$

[Notice that in a path from $(1, 1)$ to $(n, a - b)$ there are x successes and y failures, where $x + y = n - 1 = a + b - 1$, $x - y = a - b - 1$, so

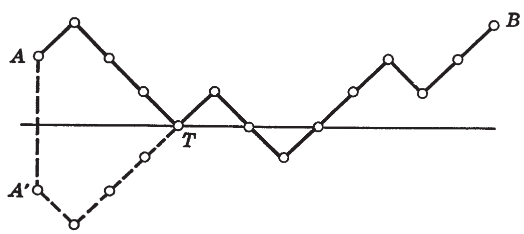

FIGURE 6.3.2 Reflection Principle.

$x = a - 1$, $y = b$. Similarly, a path from $(1, -1)$ to $(n, a - b)$ must have a successes and $b - 1$ failures.]

$$= \frac{(n-1)!}{(a-1)!\,b!} - \frac{(n-1)!}{a!\,(b-1)!} = \frac{n!}{a!\,b!}\left(\frac{a}{n} - \frac{b}{n}\right)$$

Thus

$$\text{if } a + b = n, \text{ the number of paths from } (0, 0) \text{ to } (n, a - b) \text{ lying entirely above the axis} = \left(\frac{a-b}{n}\right)\binom{n}{a} \tag{6.3.3}$$

Therefore

$$P\{S_1 > 0, \ldots, S_n > 0 \mid S_n = a - b\} = \frac{a-b}{n} = \frac{a-b}{a+b} \tag{6.3.4}$$

REMARK. This problem is equivalent to the *ballot problem:* In an election with two candidates, candidate 1 receives a votes and candidate 2 receives b votes, with $a > b$, $a + b = n$. The ballots are shuffled and counted one by one. The probability that candidate 1 will lead throughout the balloting is $(a - b)/(a + b)$. [Each possible sequence of ballots corresponds to a path from $(0, 0)$ to $(n, a - b)$; a sequence in which candidate 1 is always ahead corresponds to a path from $(0, 0)$ to $(n, a - b)$ that is always above the axis.]

We now compute

$$h_j = P\{\text{the first return to 0 occurs at time } j\}$$

Since h_j must be 0 for j odd, we may as well set $j = 2n$. Now

$$h_{2n} = P\{S_1 \neq 0, \ldots, S_{2n-1} \neq 0, S_{2n} = 0\}$$

and thus h_{2n} is the number of paths from $(0, 0)$ to $(2n, 0)$ lying above the axis, times 2 (to take into account paths lying below the axis), times $p^n q^n$, the probability of each path (see Figure 6.3.3).

The number of paths from $(0, 0)$ to $(2n, 0)$ lying above the axis is the number from $(0, 0)$ to $(2n - 1, 1)$ lying above the axis, which, by (6.3.3), is $\binom{2n-1}{a}(a - b)/(2n - 1)$ (where $a + b = 2n - 1$, $a - b = 1$, hence $a = n$,

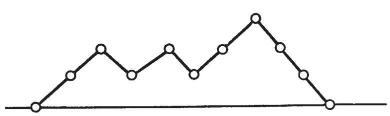

FIGURE 6.3.3 A First Return to 0 at Time $2n$.

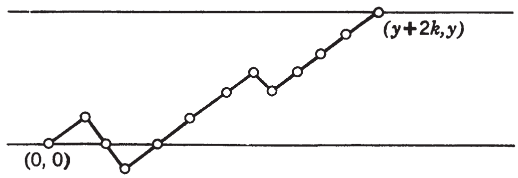

FIGURE 6.3.4 Computation of First Passage Times.

$b = n - 1$), that is,

$$\binom{2n-1}{n}\frac{1}{2n-1} = \frac{(2n-2)!}{n!\,(n-1)!} = \frac{1}{n}\binom{2n-2}{n-1}$$

Thus

$$h_{2n} = \frac{2}{n}\binom{2n-2}{n-1}(pq)^n \tag{6.3.5}$$

We now compute probabilities of *first passage times*, that is, $P\{$the first passage through $y > 0$ takes place at time $r\}$. The only possible values of r are of the form $y + 2k$, $k = 0, 1, \ldots$; hence we are looking for

$$h_{y+2k}^{(y)} = P\{\text{the first passage through } y > 0 \text{ occurs at time } y + 2k\}$$

To do the computation in an effortless manner, see Figure 6.3.4. If we look at the path of Figure 6.3.4 backward, it always lies below y and travels a vertical distance y in time $y + 2k$. Thus the number of favorable paths is the number of paths from $(0, 0)$ to $(y + 2k, y)$ that lie above the axis; that is, by (6.3.3),

$$\binom{y+2k}{a}\frac{a-b}{y+2k} \qquad \text{where } a + b = y + 2k,\ a - b = y$$

Thus $a = y + k$, $b = k$. Consequently

$$h_{y+2k}^{(y)} = \frac{y}{y+2k}\binom{y+2k}{k}p^{y+k}q^k \tag{6.3.6}$$

PROBLEMS

The first five problems refer to the simple random walk starting at 0, with no barriers.

1. Show that $P\{S_1 \geq 0, S_2 \geq 0, \ldots, S_{2n-1} \geq 0, S_{2n} = 0\} = u_{2n}/(n + 1)$, where $u_{2n} = P\{S_{2n} = 0\}$ is the probability of n successes (and n failures) in $2n$ Bernoulli trials, that is, $\binom{2n}{n}(pq)^n$.

2. Let $p = q = 1/2$.
(a) Show that $h_{2n} = u_{2n-2}/2n$, where $u_{2n} = \binom{2n}{n}(1/2)^{2n}$.
(b) Show that $u_{2n}/u_{2n-2} = 1 - 1/2n$; hence $h_{2n} = u_{2n-2} - u_{2n}$.

3. If $p = q = 1/2$, show that $P\{S_1 \neq 0, \ldots, S_{2n} \neq 0\}$, the probability of no return to the origin in the first $2n$ steps, is $u_{2n} = \binom{2n}{n}2^{-2n}$. Show also that $P\{S_1 \neq 0, \ldots, S_{2n-1} \neq 0\} = h_{2n} + u_{2n}$.

4. If $p = q = 1/2$, show that $P\{S_1 \geq 0, \ldots, S_{2n} \geq 0\} = \binom{2n}{n}2^{-2n}$.

5. If $p = q = 1/2$, show that the average length of time required to return to the origin is infinite, by using Stirling's formula to find the asymptotic expression for h_{2n}, and then showing that $\sum_{n=1}^{\infty} nh_{2n} = \infty$.

6. Two players each toss an unbiased coin independently n times. Show that the probability that each player will have the same number of heads after n tosses is $(1/2)^{2n} \sum_{k=0}^{n} \binom{n}{k}^2$.

7. By looking at Problem 6 in a slightly different way, show that $\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$.

8. A spider and a fly are situated at the corners of an n by n grid, as shown in Figure P.6.3.8. The spider walks only north or east, the fly only south or west;

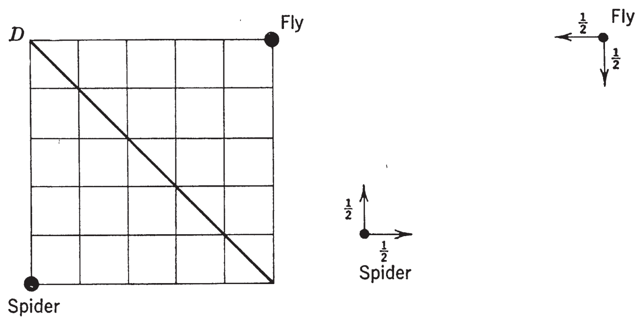

FIGURE P.6.3.8

they take their steps simultaneously, to an adjacent vertex of the grid.
(a) Show that if they meet, the point of contact must be on the diagonal D.
(b) Show that if the successive steps are independent, and equally likely to go in each of the two possible directions, the probability that they will meet is $\binom{2n}{n}(1/2)^{2n}$.

6.4 GENERATING FUNCTIONS

Let $\{a_n, n \geq 0\}$ be a bounded sequence of real numbers. The *generating function* of the sequence is defined by

$$A(z) = \sum_{n=0}^{\infty} a_n z^n, \qquad z \text{ complex}$$

The series converges at least for $|z| < 1$. If R is a discrete random variable taking on only nonnegative integer values, and $P\{R = n\} = a_n$, $n = 0, 1, \ldots$, then $A(z)$ is called the *generating function of* R. Note that $A(z) = \sum_{n=0}^{\infty} z^n P\{R = n\} = E(z^R)$, the characteristic function of R with z replacing e^{-iu}.

We have seen that the characteristic function of a sum of independent random variables is the product of the characteristic functions. An analogous result holds for generating functions.

Theorem 1. *Let $\{a_n\}$ and $\{b_n\}$ be bounded sequences of real numbers. Let $\{c_n\}$ be the convolution of $\{a_n\}$ and $\{b_n\}$, defined by*

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} \left(= \sum_{j=0}^{n} b_j a_{n-j} \right)$$

Then $C(z) = \sum_{n=0}^{\infty} c_n z^n$ is convergent at least for $|z| < 1$, and

$$C(z) = A(z)B(z)$$

PROOF. Suppose first that $a_n = P\{R_1 = n\}$, $b_n = P\{R_2 = n\}$, where R_1 and R_2 are independent nonnegative integer-valued random variables. Then $c_n = P\{R_1 + R_2 = n\}$, since $\{R_1 + R_2 = n\}$ is the disjoint union of the events $\{R_1 = k, R_2 = n - k\}$, $k = 0, 1, \ldots, n$. Thus

$$C(z) = E(z^{R_1+R_2}) = E(z^{R_1}z^{R_2}) = E(z^{R_1})E(z^{R_2}) = A(z)B(z)$$

In general,

$$\sum_{n=0}^{\infty} c_n z^n = \sum_{n=0}^{\infty} \sum_{k=0}^{n} a_k b_{n-k} z^n = \sum_{k=0}^{\infty} \left(\sum_{n=k}^{\infty} b_{n-k} z^{n-k} \right) a_k z^k = A(z)B(z)$$

We have seen that under appropriate conditions the moments of a random variable can be obtained from its characteristic function. Similar results hold for generating functions. Let $A(z)$ be the generating function of the random variable R; restrict z to be real and between 0 and 1. We show that

$$E(R) = A'(1) \tag{6.4.1}$$

where

$$A'(1) = \lim_{z \to 1} A'(z)$$

If $E(R)$ is finite, then the variance of R is given by

$$\text{Var } R = A''(1) + A'(1) - [A'(1)]^2 \tag{6.4.2}$$

To establish (6.4.1) and (6.4.2), notice that

$$A(z) = \sum_{n=0}^{\infty} a_n z^n, \qquad a_n = P\{R = n\}$$

Thus

$$A'(z) = \sum_{n=1}^{\infty} na_n z^{n-1}$$

Let $z \to 1$ to obtain

$$A'(1) = \sum_{n=1}^{\infty} na_n = E(R)$$

proving (6.4.1). Similarly,

$$A''(z) = \sum_{n=1}^{\infty} n(n-1)a_n z^{n-2}, \qquad \text{so } A''(1) = E(R^2) - E(R)$$

Therefore

$$\text{Var } R = E(R^2) - [E(R)]^2 = A''(1) + A'(1) - [A'(1)]^2$$

which is (6.4.2).

Now consider the simple random walk starting at 0, with no barriers. Let $u_n = P\{S_n = 0\}$, $h_n =$ the probability that the first return to 0 will occur at time $n = P\{S_1 \neq 0, \ldots, S_{n-1} \neq 0, S_n = 0\}$. Let

$$U(z) = \sum_{n=0}^{\infty} u_n z^n, \qquad H(z) = \sum_{n=0}^{\infty} h_n z^n$$

(For the remainder of this section, z is restricted to real values.)

If we are at the origin at time n, the first return to 0 must occur at some time k, $k = 1, 2, \ldots, n$. If the first return to 0 occurs at time k, we must be at the origin after $n - k$ additional steps. Since the events {first return to 0 at time k}, $k = 1, 2, \ldots, n$, are disjoint, we have

$$u_n = \sum_{k=1}^{n} h_k u_{n-k}, \qquad n = 1, 2, \ldots$$

Let us write this as

$$u_n = \sum_{k=0}^{n} h_k u_{n-k}, \qquad n = 1, 2, \ldots$$

This will be valid provided that we define $h_0 = 0$. Now $u_0 = 1$, since the walk starts at the origin, but $h_0 u_0 = 0$. Thus we may write

$$v_n = \sum_{k=0}^{n} h_k u_{n-k}, \qquad n = 0, 1, \ldots \tag{6.4.3}$$

where $v_n = u_n$, $n \geq 1$; $v_0 = 0 = u_0 - 1$.

Since $\{v_n\}$ is the convolution of $\{h_n\}$ and $\{u_n\}$, Theorem 1 yields

$$V(z) = H(z)U(z)$$

But

$$V(z) = \sum_{n=0}^{\infty} v_n z^n = \sum_{n=1}^{\infty} u_n z^n = U(z) - 1$$

Thus

$$U(z)(1 - H(z)) = 1 \tag{6.4.4}$$

We may use (6.4.4) to find the h_n explicitly. For

$$U(z) = \sum_{n=0}^{\infty} u_n z^n = \sum_{n=0}^{\infty} u_{2n} z^{2n} = \sum_{n=0}^{\infty} \binom{2n}{n} (pq)^n z^{2n}$$

This can be put into a closed form, as follows. We claim that

$$\binom{2n}{n} = \binom{-1/2}{n}(-4)^n \tag{6.4.5}$$

where, for any real number α, $\binom{\alpha}{n}$ denotes

$$\frac{\alpha(\alpha - 1)\cdots(\alpha - n + 1)}{n!}$$

To see this, write

$$\begin{aligned}
\binom{-1/2}{n} &= \frac{(-1/2)(-3/2)\cdots[-(2n-1)/2]}{n!} \\
&= (-1)^n \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{n!\, 2^n} \frac{n!}{n!} \\
&= \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{n!} \frac{(2/2)(4/2)(6/2)\cdots(2n/2)}{n!\, 2^n} (-1)^n \\
&= \frac{(2n)!}{n!\, n!} \frac{(-1)^n}{4^n}
\end{aligned}$$

proving (6.4.5). Thus

$$U(z) = \sum_{n=0}^{\infty} \binom{-1/2}{n} (-4pqz^2)^n = (1 - 4pqz^2)^{-1/2}$$

by the binomial theorem. By (6.4.4) we have

$$H(z) = 1 - \frac{1}{U(z)} = 1 - (1 - 4pqz^2)^{1/2} \tag{6.4.6}$$

This may be expanded by the binomial theorem to obtain the h_n (see Problem 1); of course the results will agree with (6.3.5), obtained by the combinatorial approach of the preceding section. Notice that we must have the positive square root in (6.4.6), since $H(0) = h_0 = 0$.

Some useful information may be gathered without expanding $H(z)$. Observe that $H(1) = \sum_{n=0}^{\infty} h_n$ is the probability of eventual return to 0.

By (6.4.6),

$$\begin{aligned} H(1) &= 1 - (1 - 4pq)^{1/2} \\ &= 1 - |p - q| \qquad \text{by (6.2.2)} \end{aligned}$$

This agrees with the result (6.2.7) obtained previously.

Now assume $p = q = 1/2$, so that there is a return to 0 with probability 1. We show that the average length of time required to return to the origin is infinite. For if T is the time of first return to 0, then

$$E(T) = \sum_{n=1}^{\infty} nP\{T = n\} = \sum_{n=1}^{\infty} nh_n = H'(1)$$

as in (6.4.1). By (6.4.6),

$$H'(z) = -\frac{d}{dz}(1 - z^2)^{1/2} = z(1 - z^2)^{-1/2} \to \infty \qquad \text{as } z \to 1$$

Thus $E(T) = \infty$, as asserted (see Problem 5, Section 6.3, for another approach).

PROBLEMS

1. Expand (6.4.6) by the binomial theorem to obtain the h_n.
2. Solve the difference equation $a_{n+1} - 3a_n = 4$ by taking the generating function of both sides to obtain
$$A(z) = \frac{4z}{(1 - z)(1 - 3z)} + \frac{a_0}{1 - 3z}$$
Expand in partial fractions and use a geometric series expansion to find a_n.
3. Let $A(z)$ be the generating function of the sequence $\{a_n\}$; assume that $\sum_{n=0}^{\infty} |a_n - a_{n-1}| < \infty$. Show that if $\lim_{n\to\infty} a_n$ exists, the limit is
$$\lim_{z\to 1} (1 - z)A(z)$$
4. If R is a random variable with generating function $A(z)$, find the generating function of $R + k$ and kR, where k is a nonnegative integer. If $F(n) = P\{R \le n\}$, find the generating function of $\{F(n)\}$.
5. Let $R_1, R_2, \ldots$ be independent random variables, with $P\{R_i = 1\} = p$, $P\{R_i = 0\} = q = 1 - p$, $i = 1, 2, \ldots$. Thus we have an infinite sequence of Bernoulli trials; $R_i = 1$ corresponds to a success on trial i, and $R_i = 0$ is a failure. (Assume $0 < p < 1$.) Let R be the number of trials required to obtain the first success.
   (a) Show that $P\{R = k\} = q^{k-1}p$, $k = 1, 2, \ldots$.
   (b) Use generalized characteristic functions to show that $E(R) = 1/p$, $\operatorname{Var} R = (1 - p)/p^2$; check the result by calculating the generating function of R and using (6.4.1) and (6.4.2). R is said to have the *geometric* distribution.

6. With the R_i as in Problem 5, let N_r be the number of trials required to obtain the rth success ($r = 1, 2, \ldots$).

(a) Show that $P\{N_r = k\} = \binom{k-1}{r-1}p^r q^{k-r}$, $k = r, r+1, \ldots$

$$= \binom{-r}{k-r}p^r(-q)^{k-r},\ k = r, r+1, \ldots$$

where $\binom{-r}{j}$ is defined as $(-r)(-r-1)\cdots(-r-j+1)/j!$, $j = 1, 2, \ldots,$ $\binom{-r}{0} = 1.$

(b) Let T_1 = the number of trials required to obtain the first success, T_2 = the number of trials following the first success up to and including the second success, ..., T_r = the number of trials following the $(r-1)$st success up to and including the rth success (thus $N_r = T_1 + \cdots + T_r$). Show that the T_i are independent, each with the geometric distribution.

(c) Show that $E(N_r) = r/p$, $\operatorname{Var} N_r = r(1-p)/p^2$. Find the characteristic function and the generating function of N_r. N_r is said to have the *negative binomial* distribution.

7. With the R_i as in Problem 5, let R be the length of the run (of either successes or failures) started by the first trial. Find $P\{R = k\}$, $k = 1, 2, \ldots,$ and $E(R)$.

8. In Problem 6, find the joint probability functions of N_1 and N_2; also find (in a relatively effortless manner) the correlation coefficient between N_1 and N_2.

6.5 THE POISSON RANDOM PROCESS

We now consider a mathematical model that fits a wide variety of physical phenomena. Let $T_1, T_2, \ldots$ by a sequence of independent random variables, where each T_i is absolutely continuous with density $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$; $f(x) = 0$, $x < 0$ (λ is a fixed positive constant). Let $A_n = T_1 + \cdots + T_n$, $n = 1, 2, \ldots$. We may think of A_n as the arrival time of the nth customer at a serving counter, so that T_n is the waiting time between the arrival of the $(n-1)$st customer and the arrival of the nth customer. Equally well, A_n may be regarded as the time at which the nth call is made at a telephone exchange, the time at which the nth component fails on an assembly line, or the time at which the nth electron arrives at the anode of a vacuum tube.

If $t \geq 0$, let R_t be the number of customers that have arrived up to and including time t; that is, $R_t = n$ if $A_n \leq t < A_{n+1}$ ($n = 0, 1, \ldots$; define $A_0 = 0$). A sketch of $(R_t, t \geq 0)$ is given in Figure 6.5.1.

Thus we have a *family* of random variables R_t, $t \geq 0$ (not just a sequence, but instead a random variable defined for each nonnegative real number). A family of random variables R_t, where t ranges over an arbitrary set I, is called a *random process* or *stochastic process*. Note that if I is the set of positive integers, the random process becomes a sequence of random variables; if I is a finite set, we obtain a finite collection of random variables; and if I consists of only one element, we obtain a single random variable. Thus the concept of a random process includes all situations studied previously.

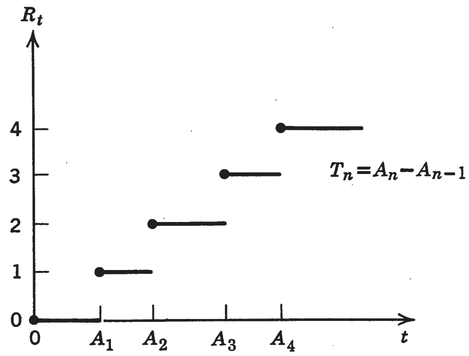

FIGURE 6.5.1 Poisson Process.

If the outcome of the experiment is ω, we may regard $(R_t(\omega),\ t \in I)$ as a real-valued function defined on I. In Figure 6.5.1 what is actually sketched is $R_t(\omega)$ versus t, $t \in I =$ the nonnegative reals, for a particular ω. Thus, roughly speaking, we have a "random function," that is, a function that depends on the outcome of a random experiment.

The particular process introduced above is called the *Poisson process* since, for each $t > 0$, R_t has the Poisson distribution with parameter λt. Let us verify this.

If k is a nonnegative integer,

$$\begin{aligned} P\{R_t \le k\} &= P\{\text{at most } k \text{ customers have arrived by time } t\} \\ &= P\{(k+1)\text{st customer arrives after time } t\} \\ &= P\{T_1 + \cdots + T_{k+1} > t\} \end{aligned}$$

But $A_{k+1} = T_1 + \cdots + T_{k+1}$ is the sum of $k + 1$ independent random variables, each with generalized characteristic function $\int_0^\infty \lambda e^{-\lambda x} e^{-sx}\,dx = \lambda/(s + \lambda)$, Re $s > -\lambda$; hence A_{k+1} has the generalized characteristic function $[\lambda/(s + \lambda)]^{k+1}$, Re $s > -\lambda$. Thus (see Table 5.1.1) the density of A_{k+1} is

$$f_{A_{k+1}}(x) = \frac{1}{k!}\,\lambda^{k+1} x^k e^{-\lambda x} u(x) \tag{6.5.1}$$

where u is the unit step function. Thus

$$\begin{aligned} P\{R_t \le k\} &= P\{T_1 + \cdots + T_{k+1} > t\} \\ &= \int_t^\infty \frac{1}{k!}\,\lambda^{k+1} x^k e^{-\lambda x}\,dx \\ &= \int_t^\infty \frac{1}{k!}\,\lambda^{k+1} x^k\, d\left(\frac{-e^{-\lambda x}}{\lambda}\right) \\ &= \frac{1}{k!}\,(\lambda t)^k e^{-\lambda t} + \int_t^\infty \frac{1}{(k-1)!}\,\lambda^k x^{k-1} e^{-\lambda x}\,dx \quad \text{(integrate by parts)} \end{aligned}$$

Integrating by parts successively, we obtain

$$P\{R_t \le k\} = \sum_{i=0}^{k} e^{-\lambda t} \frac{(\lambda t)^i}{i!}$$

Hence R_t has the Poisson distribution with parameter λt.

Now the mean of a Poisson random variable is its parameter, so that $E(R_t) = \lambda t$. Thus λ may be interpreted as the average number of customers arriving per second. We should expect that λ^{-1} is the average number of seconds per customer, that is, the average waiting time between customers. This may be verified by computing $E(T_i)$.

$$E(T_i) = \int_0^\infty \lambda x e^{-\lambda x}\, dx = \frac{1}{\lambda}$$

We now establish an important feature of the Poisson process. Intuitively, if we arrive at the serving counter at time t and the last customer to arrive came at time $t - h$, the distribution function of the length of time we must wait for the arrival of the next customer does not depend on h, and in fact coincides with the distribution function of T_1. Thus we are essentially starting from scratch at time t; the process does not remember that h seconds have elapsed between the arrival of the last customer and the present time.

If W_t is the waiting time from t to the arrival of the next customer, we wish to show that $P\{W_t \le z\} = P\{T_1 \le z\}$, $z \ge 0$. We have

$$\begin{aligned} P\{W_t \le z\} &= P\{\text{for some } n = 1, 2, \ldots, \text{ the } n\text{th customer arrives in} \\ &\qquad (t, t+z] \text{ and the } (n+1)\text{st customer arrives after time } t+z\} \\ &= P\left[\bigcup_{n=1}^{\infty} \{t < A_n \le t + z < A_{n+1}\}\right] \qquad (6.5.2) \end{aligned}$$

(see Figure 6.5.2). To justify (6.5.2), notice that if $t < A_n \le t + z < A_{n+1}$ for some n, then $W_t \le z$. Conversely, if $W_t \le z$, then some customer arrives in $(t, t + z]$ and hence there will be a last customer to arrive in the interval. (If not, then $\sum_{n=1}^\infty T_n < \infty$; but this event has probability 0; see Problem 1.) Now $P\{t < A_n \le t + z < A_{n+1}\} = P\{t < A_n \le t + z, A_n + T_{n+1} > t + z\}$.

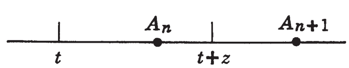

FIGURE 6.5.2

Since $A_n(= T_1 + \cdots + T_n)$ and T_{n+1} are independent, we obtain, by (6.5.1),

$$
\begin{aligned}
P\{t < A_n \le t + z < A_{n+1}\} &= \iint\limits_{\substack{t < x \le t+z,\\ x+y > t+z}} \frac{1}{(n-1)!} \lambda^n x^{n-1} e^{-\lambda x} \lambda e^{-\lambda y}\, dx\, dy \\
&= \frac{\lambda^n}{(n-1)!} \int_t^{t+z} x^{n-1} e^{-\lambda x} \int_{t+z-x}^{\infty} \lambda e^{-\lambda y}\, dy\, dx \\
&= \frac{\lambda^n}{(n-1)!} \int_t^{t+z} x^{n-1} e^{-\lambda x} e^{-\lambda(t+z-x)}\, dx \\
&= \frac{1}{n!} \lambda^n e^{-\lambda(t+z)}[(t+z)^n - t^n]
\end{aligned}
$$

Since $\sum_{n=0}^{\infty} r^n/n! = e^r$, (6.5.2) yields

$$
\begin{aligned}
P\{W_t \le z\} &= e^{-\lambda(t+z)}[e^{\lambda(t+z)} - 1 - (e^{\lambda t} - 1)] \\
&= 1 - e^{-\lambda z}
\end{aligned}
$$

Thus W_t has the same distribution function as T_1.

Alternatively, we may write

$$
P\{W_t \le z\} = P\left[\bigcup_{n=0}^{\infty} \{A_n \le t < A_{n+1} \le t + z\}\right]
$$

For if $A_n \le t < A_{n+1} \le t + z$ for some n, then $W_t \le z$. Conversely, if $W_t \le z$, then some customer arrives in $(t, t + z]$, and there must be a first customer to arrive in this interval, say customer $n + 1$. Thus $A_n \le t < A_{n+1} \le t + z$ for some $n \ge 0$. An argument very similar to the above shows that

$$
P\{A_n \le t < A_{n+1} \le t + z\} = \frac{(\lambda t)^n}{n!}(e^{-\lambda t} - e^{-\lambda(t+z)})
$$

and therefore $P\{W_t \le z\} = 1 - e^{-\lambda z}$ as before. In this approach, we do not have the problem of showing that $P\left\{\sum_{n=1}^{\infty} T_n < \infty\right\} = 0$.

To justify completely the statement that the process starts from scratch at time t, we may show that if $V_1, V_2, \ldots$ are the successive waiting times starting at t (so $V_1 = W_t$), then $V_1, V_2, \ldots$ are independent, and V_i and T_i have the same distribution function for all i. To see this, observe that

$$
\begin{aligned}
&P\{V_1 \le x_1, \ldots, V_k \le x_k\} \\
&\qquad = P\left[\bigcup_{n=0}^{\infty} \{A_n \le t < A_{n+1} \le t + x_1, T_{n+2} \le x_2, \ldots, T_{n+k} \le x_k\}\right]
\end{aligned}
$$

For it is clear that the set on the right is a subset of the set on the left. Conversely, if $V_1 \le x_1, \ldots, V_k \le x_k$, then a customer arrives in $(t, t + x_1]$, and hence there is a first customer in this interval, say customer $n + 1$. Then $A_n \le t < A_{n+1} \le t + x_1$, and also $V_i = T_{n+i}$, $i = 2, \ldots, k$, as desired. Therefore

$$\begin{aligned}
P\{V_1 \le x_1, \ldots, V_k \le x_k\} &= \left[\sum_{n=0}^{\infty} P\{A_n \le t < A_{n+1} \le t + x_1\}\right] \\
&\quad \times \prod_{i=2}^{k} P\{T_{n+i} \le x_i\} \\
&= P\{W_t \le x_1\} \prod_{i=2}^{k} P\{T_i \le x_i\} \\
&= \prod_{i=1}^{k} P\{T_i \le x_i\}
\end{aligned}$$

Fix j and let $x_i \to \infty$, $i \ne j$, to conclude that $P\{V_j \le x_j\} = P\{T_j \le x_j\}$. Consequently

$$P\{V_1 \le x_1, \ldots, V_k \le x_k\} = \prod_{i=1}^{k} P\{V_i \le x_i\}$$

and the result follows. In particular, the number of customers arriving in the interval $(t, t + \tau]$ has the Poisson distribution with parameter $\lambda\tau$.

The "memoryless" feature of the Poisson process is connected with a basic property of the exponential density, as the following intuitive argument shows. Suppose that we start counting at time t, and the last customer to arrive, say customer $n - 1$, came at time $t - h$. Then (see Figure 6.5.3) the probability that $W_t \le z$ is the probability that $T_n \le z + h$, given that we have already waited h seconds; that is, given that $T_n > h$. Thus

$$\begin{aligned}
P\{W_t \le z\} &= P\{T_n \le z + h \mid T_n > h\} \\
&= \frac{P\{h < T_n \le z + h\}}{P\{T_n > h\}} \\
&= \frac{\displaystyle\int_h^{z+h} \lambda e^{-\lambda x}\, dx}{\displaystyle\int_h^{\infty} \lambda e^{-\lambda x}\, dx} \\
&= \frac{e^{-\lambda h} - e^{-\lambda(z+h)}}{e^{-\lambda h}} = 1 - e^{-\lambda z} = P\{T_n \le z\}
\end{aligned}$$

so that W_t and T_n have the same distribution function (for any n).

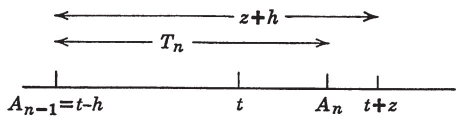

FIGURE 6.5.3

The key property of the exponential density used in this argument is

$$P\{T_n \le z + h \mid T_n > h\} = P\{T_n \le z\}$$

that is,

$$P\{T_n > z + h \mid T_n > h\} = P\{T_n > z\}$$

or (since $\{T_n > h,\, T_n > z + h\} = \{T_n > z + h\}$)

$$P\{T_n > z + h\} = P\{T_n > z\}P\{T_n > h\}$$

In fact, a positive random variable T having the property that

$$P\{T > z + h\} = P\{T > z\}P\{T > h\} \qquad \text{for all } z,\, h \ge 0$$

must have exponential density. For if $G(z) = P\{T > z\}$, then

$$G(z + h) = G(z)G(h)$$

Therefore

$$\frac{G(z + h) - G(z)}{h} = G(z)\left(\frac{G(h) - 1}{h}\right) = G(z)\left(\frac{G(h) - G(0)}{h}\right) \quad \text{since } G(0) = 1$$

Let $h \to 0$ to conclude that

$$G'(z) = G'(0)G(z)$$

This is a differential equation of the form

$$\frac{dy}{dx} + \lambda y = 0 \qquad (\lambda = -G'(0))$$

whose solution is $y = ce^{-\lambda x}$; that is,

$$G(z) = ce^{-\lambda z}$$

But $G(0) = c = 1$; hence the distribution function of T is

$$F_T(z) = 1 - G(z) = 1 - e^{-\lambda z}, \qquad z \ge 0$$

and thus the density of T is

$$f_T(z) = \lambda e^{-\lambda z}, \qquad z \geq 0$$

as desired.

(The above argument assumes that T has a continuous density, but actually the result is true without this requirement.)

The memoryless feature of the Poisson process may be used to show that the process satisfies the *Markov property*:

If $0 \leq t_1 < t_2 < \cdots < t_{n+1}$ and $a_1, \ldots, a_{n+1}$ are nonnegative integers with $a_1 \leq \cdots \leq a_{n+1}$, then

$$P\{R_{t_{n+1}} = a_{n+1} \mid R_{t_1} = a_1, \ldots, R_{t_n} = a_n\} = P\{R_{t_{n+1}} = a_{n+1} \mid R_{t_n} = a_n\}$$

Thus the behavior of the process at the "future" time t_{n+1}, given the behavior at "past" times $t_1, \ldots, t_{n-1}$ and the "present" time t_n, depends only on the present state, or the number of customers at time t_n. For example,

$$\begin{aligned} P\{R_{t_4} = 15 \mid R_{t_1} = 0, R_{t_2} = 3, R_{t_3} = 8\} &= P\{R_{t_4} = 15 \mid R_{t_3} = 8\} \\ &= \text{the probability that exactly 7 customers will arrive between } t_3 \text{ and } t_4 \\ &= e^{-\lambda(t_4 - t_3)} \frac{[\lambda(t_4 - t_3)]^7}{7!} \end{aligned}$$

This result is reasonable in view of the memoryless feature, but a formal proof becomes quite cumbersome and will be omitted.

We consider the Markov property in detail in the next chapter.

We close this section by describing a physical approach to the Poisson process. Suppose that we divide the interval $(t, t + \tau]$ into subintervals of length Δt, and assume that the subintervals are so small that the probability of the arrival of more than one customer in a given subinterval I is negligible. If the average number of customers arriving per second is λ, and R_I is the number of customers arriving in the subinterval I, then $E(R_I) = \lambda\,\Delta t$. But $E(R_I) = (0)P\{R_I = 0\} + (1)P\{R_I = 1\} = P\{R_I = 1\}$, so that with probability $\lambda\,\Delta t$ a customer will arrive in I, and with probability $1 - \lambda\,\Delta t$ no customer will arrive.

We assume that if I and J are disjoint subintervals, R_I and R_J are independent. Then we have a sequence of $n = \tau/\Delta t$ Bernoulli trials, with probability $p = \lambda\,\Delta t$ of success on a given trial, and with Δt very small. Thus we expect that the number $N(t, t + \tau]$ of customers arriving in $(t, t + \tau]$ should have the Poisson distribution with parameter $\lambda\tau$. Furthermore, if W_t is the waiting time from t to the arrival of the next customer, then $P\{W_t > x\} = P\{N(t, t + x] = 0\} = e^{-\lambda x}$, so that W_t has exponential density.

PROBLEMS

1. Show that $P\{\sum_{n=1}^{\infty} T_n < \infty\} = 0$.

2. Show that the probability that an even number of customers will arrive in the interval $(t, t + \tau]$ is $(1/2)(1 + e^{-2\lambda\tau})$ and the probability that an odd number of customers will arrive in this interval is $(1/2)(1 - e^{-2\lambda\tau})$.

3. (The random telegraph signal) Let $T_1, T_2, \ldots$ be independent, each with density $\lambda e^{-\lambda x}u(x)$. Define a random process R_t, $t \geq 0$, as follows.
$R_0 = +1$ or -1 with equal probability (assume R_0 independent of the T_i)

$$R_t = R_0,\ 0 \leq t < T_1$$

$$R_t = -R_0,\ T_1 \leq t < T_1 + T_2$$

$$R_t = R_0,\ T_1 + T_2 \leq t < T_1 + T_2 + T_3$$

$$\vdots$$

$$R_t = (-1)^n R_0,\ T_1 + \cdots + T_n \leq t < T_1 + \cdots + T_{n+1}$$

(see Figure P.6.5.3).

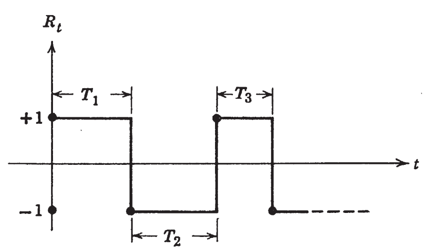

FIGURE P.6.5.3

(a) Find the joint probability function of R_t and $R_{t+\tau}$ $(t, \tau \geq 0)$.
(b) Find the *covariance function* of the process, defined by $K(t, \tau) = \text{Cov}(R_t, R_{t+\tau})$, $t, \tau \geq 0$.

*6.6 THE STRONG LAW OF LARGE NUMBERS

In this section we show that the arithmetic average $(R_1 + \cdots + R_n)/n$ of a sequence of independent observations of a random variable R converges with probability 1 to $E(R)$ (assumed finite). In other words, we have

$$\lim_{n\to\infty} \frac{R_1(\omega) + \cdots + R_n(\omega)}{n} = E(R)$$

for all ω, except possibly on a set of probability 0. We shall see that this is a stronger convergence statement than the weak law of large numbers.

Let $(\Omega, \mathscr{F}, P)$ be a probability space, fixed throughout the discussion. If $A_1, A_2, \ldots$ is a sequence of events, we define the *upper limit* or *limit superior* of the sequence as

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \tag{6.6.1}$$

and the *lower limit* or *limit inferior* of the sequence as

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \tag{6.6.2}$$

Theorem 1. $\limsup_n A_n = \{\omega\colon \omega \in A_n$ *for infinitely many* $n\}$, $\liminf_n A_n = \{\omega\colon \omega \in A_n$ *eventually, i.e., for all but finitely many* $n\}$.

Proof. By (6.6.1), $\omega \in \limsup A_n$ iff, for every n, $\omega \in \bigcup_{k=n}^{\infty} A_k$; that is, for every n there is a $k \geq n$ such that $\omega \in A_k$, or $\omega \in A_n$ for infinitely many n. By (6.6.2), $\omega \in \liminf_n A_n$ iff, for some n, $\omega \in \bigcap_{k=n}^{\infty} A_k$; that is, for some n, $\omega \in A_k$ for all $k \geq n$, or $\omega \in A_n$ eventually.

Theorem 2. *Let R, R_1, $R_2, \ldots$ be random variables on $(\Omega, \mathscr{F}, P)$. Denote by $\{R_n \to R\}$ the set $\{\omega\colon \lim_{n\to\infty} R_n(\omega) = R(\omega)\}$. Then $\{R_n \to R\} = \bigcap_{m=1}^{\infty} \liminf_n A_{nm}$, where $A_{nm} = \{\omega\colon |R_n(\omega) - R(\omega)| < 1/m\}$.*

Proof. $R_n(\omega) \to R(\omega)$ iff for every $m = 1, 2, \ldots, |R_n(\omega) - R(\omega)| < 1/m$ eventually, that is (Theorem 1), for every $m = 1, 2, \ldots\ \omega \in \liminf_n A_{nm}$.

We say that the sequence $R_1, R_2, \ldots$ converges *almost surely* to R (notation: $R_n \xrightarrow{\text{a.s.}} R$) iff $P\{R_n \to R\} = 1$. The terminology "almost everywhere" is also used.

Theorem 3. *$R_n \xrightarrow{\text{a.s.}} R$ iff for every $\varepsilon > 0$, $P\{|R_k - R| \geq \varepsilon$ for at least one $k \geq n\} \to 0$ as $n \to \infty$.*

Proof. By Theorem 2, $P\{R_n \to R\} \leq P\{\liminf_n A_{nm}\}$ for every m; hence $R_n \xrightarrow{\text{a.s.}} R$ implies that $\liminf_n A_{nm}$ has probability 1 for every m. Conversely, if $P(\liminf_n A_{nm}) = 1$ for all m, then, by Theorem 2, $\{R_n \to R\}$

is a countable intersection of sets with probability 1 and hence has probability 1 (see Problem 1).

Now in (6.6.2), $\bigcap_{k=n}^{\infty} A_k$, $n = 1, 2, \ldots$ is an expanding sequence; hence $P(\liminf_n A_n) = \lim_{n\to\infty} P(\bigcap_{k=n}^{\infty} A_k)$. Thus

$$
\begin{aligned}
R_n \xrightarrow{\text{a.s.}} R \quad &\text{iff} \quad P(\lim_n \inf A_{nm}) = 1 \qquad \text{for all } m = 1, 2, \ldots \\
&\text{iff} \quad \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty} A_{km}\right) = 1 \qquad \text{for all } m = 1, 2, \ldots \\
&\text{iff} \quad P\left\{|R_k - R| < \frac{1}{m} \quad \text{for all } k \geq n\right\} \to 1 \text{ as } n \to \infty \\
&\qquad\qquad \text{for all } m = 1, 2, \ldots \\
&\text{iff} \quad P\left\{|R_k - R| \geq \frac{1}{m} \quad \text{for at least one } k \geq n\right\} \to 0 \text{ as } n \to \infty \\
&\qquad\qquad \text{for all } m = 1, 2, \ldots \\
&\text{iff} \quad P\{|R_k - R| \geq \varepsilon \quad \text{for at least one } k \geq n\} \to 0 \text{ as } n \to \infty \\
&\qquad\qquad \text{for all } \varepsilon > 0
\end{aligned}
$$

(see Problem 2).

COROLLARY. $R_n \xrightarrow{\text{a.s.}} R$ implies $R_n \xrightarrow{P} R$.

PROOF. $R_n \xrightarrow{P} R$ iff for every $\varepsilon > 0$, $P\{|R_n - R| \geq \varepsilon\} \to 0$ as $n \to \infty$ (see Section 5.4). Now $\{|R_k - R| \geq \varepsilon$ for at least one $k \geq n\} \supset \{|R_n - R| \geq \varepsilon\}$, so that $P\{|R_k - R| \geq \varepsilon$ for at least one $k \geq n\} \geq P\{|R_n - R| \geq \varepsilon\}$. The result now follows from Theorem 3.

For an example in which $R_n \xrightarrow{P} R$ but $R_n \overset{\text{a.s.}}{\not\to} R$, see Problem 3.

Theorem 4 (Borel-Cantelli Lemma). *If $A_1, A_2, \ldots$ are events in a given probability space, and $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(\limsup_n A_n) = 0$; that is, the probability that A_n occurs for infinitely many n is 0.*

PROOF. By (6.6.1),

$$
\begin{aligned}
P\left(\limsup_n A_n\right) &\leq P\left(\bigcup_{k=n}^{\infty} A_k\right) \qquad \text{for every } n \\
&\leq \sum_{k=n}^{\infty} P(A_k) \qquad \text{by (1.3.10)} \\
&= \sum_{k=1}^{\infty} P(A_k) - \sum_{k=1}^{n-1} P(A_k) \to 0 \qquad \text{as } n \to \infty
\end{aligned}
$$

since $\sum_n P(A_n) < \infty$.

Theorem 5. *If for every* $\varepsilon > 0$, $\sum_{n=1}^{\infty} P\{|R_n - R| \geq \varepsilon\} < \infty$, *then* $R_n \xrightarrow{\text{a.s.}} R$.

PROOF. By Theorem 4, $\limsup_n \{|R_n - R| \geq \varepsilon\}$ has probability 0 for every $\varepsilon > 0$; that is, the probability that $|R_n - R| \geq \varepsilon$ for infinitely many n is 0. Take complements to show that the probability that $|R_n - R| \geq \varepsilon$ for only finitely many n is 1; that is, with probability 1, $|R_n - R| < \varepsilon$ eventually. Since ε is arbitrary, we may set $\varepsilon = 1/m$, $m = 1, 2, \ldots$. Thus $P(\liminf_n A_{nm}) = 1$ for $m = 1, 2, \ldots$, and the result follows by Theorem 2.

Theorem 6. *If* $\sum_{n=1}^{\infty} E[|R_n - R|^k] < \infty$ *for some* $k > 0$, *then* $R_n \xrightarrow{\text{a.s.}} R$.

PROOF. $P\{|R_n - R| \geq \varepsilon\} \leq E[|R_n - R|^k]/\varepsilon^k$ by Chebyshev's inequality, and the result follows by Theorem 5.

Theorem 7 (Strong Law of Large Numbers). *Let* $R_1, R_2, \ldots$ *be independent random variables on a given probability space. Assume uniformly bounded fourth central moments; that is, for some positive real number* M,

$$E[|R_i - E(R_i)|^4] \leq M$$

for all i. *Let* $S_n = R_1 + \cdots + R_n$. *Then*

$$\frac{S_n - E(S_n)}{n} \xrightarrow{\text{a.s.}} 0$$

In particular, if $E(R_i) = m$ *for all* i, *then* $E(S_n) = nm$; *hence*

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} m$$

PROOF. Since $S_n - E(S_n) = \sum_{i=1}^{n} R_i'$, where

$$R_i' = R_i - E(R_i), \qquad E(R_i') = 0, \qquad E(|R_i'|^4) = E[|R_i - E(R_i)|^4] \leq M < \infty$$

we may assume without loss of generality that all $E(R_i)$ are 0. We show that $\sum_{n=1}^{\infty} E[(S_n/n)^4] < \infty$, and the result will follow by Theorem 6. Now

$$\begin{aligned} S_n^4 &= (R_1 + \cdots + R_n)^4 \\ &= \sum_{j=1}^{n} R_j^4 + \sum_{\substack{j,k=1 \\ j<k}}^{n} \frac{4!\, R_j^2 R_k^2}{2!\,2!} + \sum_{\substack{j \neq k \\ j \neq l \\ k<l}} \frac{4!}{2!\,1!\,1!} R_j^2 R_k R_l + \sum_{j<k<l<m} 4!\, R_j R_k R_l R_m \\ &\quad + \sum_{j \neq k} \frac{4!}{3!\,1!} R_j^3 R_k \end{aligned}$$

But

$$E(R_j^2 R_k R_l) = E(R_j^2)E(R_k)E(R_l) \qquad \text{by independence}$$
$$= 0$$

Similarly,

$$E(R_j R_k R_l R_m) = E(R_j^3 R_k) = 0$$

Thus

$$E(S_n^4) = \sum_{j=1}^{n} E(R_j^4) + \sum_{\substack{j,k=1 \\ j<k}}^{n} 6E(R_j^2)E(R_k^2)$$

By the Schwarz inequality,

$$E(R_j^2) = E(R_j^2 \cdot 1) \le [E(R_j^4)E(1^2)]^{1/2} \le M^{1/2}$$

Hence

$$E(S_n^4) \le nM + 6\,\frac{n(n-1)}{2}\,M = (3n^2 - 2n)M < 3n^2 M$$

Consequently

$$\sum_{n=1}^{\infty} \frac{E(S_n^4)}{n^4} < \sum_{n=1}^{\infty} \frac{3n^2 M}{n^4} = 3M \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

The theorem is proved.

REMARKS. If all R_i have the same distribution function, it turns out that the hypothesis on the fourth central moments can be replaced by the assumption that the $E(R_i)$ are finite [of course, in this case $E(R_i)$ is the same for all i]. In general, the hypothesis on the fourth central moments can be replaced by the assumption that for some M and $\delta > 0$, $E[|R_i - E(R_i)|^{1+\delta}] \le M$ for all i.

Now consider an infinite sequence of Bernoulli trials, with $R_i = 1$ if there is a success on trial i, $R_i = 0$ if there is a failure on trial i. Then

$$\frac{S_n}{n} = \frac{R_1 + \cdots + R_n}{n}$$

is the relative frequency of successes in n trials and $E(R_i)$ is the probability p of success on a given trial. The strong law of large numbers says that if we regard the observation of all the R_i as one performance of the experiment, the relative frequency of successes will almost certainly converge to p. The weak law of large numbers says only that if we consider a sufficiently large but fixed n (essentially regarding observation of $R_1, \ldots, R_n$ as one performance of the experiment), the probability that the relative frequency will be within a specified distance ε of p is $> 1 - \delta$, where $\delta > 0$ is preassigned. The requirements on n will depend on ε and δ, as well as p. Recall that, by

Chebyshev's inequality,

$$P\left\{\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right\} \leq \frac{E[(S_n - ES_n)^2/n^2]}{\varepsilon^2} = \frac{\text{Var } S_n}{n^2\varepsilon^2} = \frac{\sum_{j=1}^{n} \text{Var } R_j}{n^2\varepsilon^2}$$
$$= \frac{p(1-p)}{n\varepsilon^2}$$

Thus

$$P\left\{\left|\frac{S_n}{n} - p\right| < \varepsilon\right\} \geq 1 - \frac{p(1-p)}{n\varepsilon^2} > 1 - \delta \qquad \text{for large enough } n$$

PROBLEMS

1. Show that a countable intersection of sets with probability 1 still has probability 1. Does this hold for an uncountable intersection?
2. If $P\{|R_k - R| \geq \varepsilon$ for at least one $k \geq n\} \to 0$ as $n \to \infty$ for all ε of the form $1/m$, $m = 1, 2, \ldots$, show that this holds for all $\varepsilon > 0$.
3. Let R_0 be uniformly distributed on the interval $(0, 1]$. Define the following sequence of random variables.

$$\begin{aligned}
&R_1 = g_1(R_0) \equiv 1 \\
&R_{21} = g_{21}(R_0) = 1 \quad \text{if } 0 < R_0 \leq \tfrac{1}{2}; \qquad R_{21} = 0 \text{ otherwise} \\
&R_{22} = 1 \quad \text{if } \tfrac{1}{2} < R_0 \leq 1, \quad 0 \text{ otherwise} \\
&R_{31} = 1 \quad \text{if } 0 < R_0 \leq \tfrac{1}{3}, \quad 0 \text{ otherwise} \\
&R_{32} = 1 \quad \text{if } \tfrac{1}{3} < R_0 \leq \tfrac{2}{3}, \quad 0 \text{ otherwise} \\
&R_{33} = 1 \quad \text{if } \tfrac{2}{3} < R_0 \leq 1, \quad 0 \text{ otherwise}
\end{aligned}$$

In general, let

$$R_{nm} = 1 \quad \text{if } \frac{m-1}{n} < R_0 \leq \frac{m}{n}, \qquad n = 1, 2, \ldots, m = 1, 2, \ldots, n,$$
$$0 \text{ otherwise}$$

(see Figure P.6.6.3 for $n = 4$).

The fact that we are using two subscripts is unimportant. We may arrange the R_{nm} as a single sequence.

$$R_1, R_{21}, R_{22}, R_{31}, R_{32}, R_{33}, R_{41}, R_{42}, R_{43}, R_{44}, \qquad \text{etc.}$$

Show that the sequence converges in probability to 0, but does not converge almost surely to 0. In fact, $P\{R_{nm} \to 0\} = 0$.

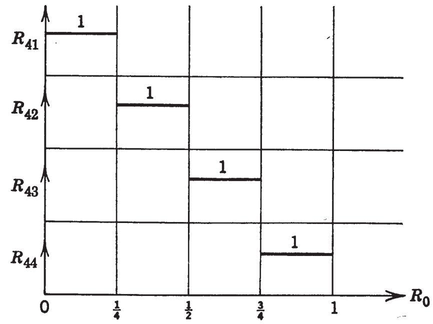

FIGURE P.6.6.3

4. Let $A_1, A_2, \ldots$ be an arbitrary sequence of events in a given probability space.
(a) Show that $\liminf_n A_n \subset \limsup_n A_n$.
(b) If the A_n form an expanding sequence whose union is A, or a contracting sequence whose intersection is A, show that $\liminf_n A_n = \limsup_n A_n = A$.
(c) In general, if $\limsup_n A_n = \liminf_n A_n = A$, we say that A is the *limit* of the sequence $\{A_n\}$ (notation: $A = \lim_n A_n$). Give an example of a sequence that is not eventually expanding or contracting (i.e., that does not become an expanding or contracting sequence if an appropriate finite number of terms is omitted), but that has a limit.
(d) If $A = \lim_n A_n$, show that $P(A) = \lim_{n\to\infty} P(A_n)$. HINT: $\bigcap_{k=n}^{\infty} A_k$ expands to $\liminf A_n$, $\bigcup_{k=n}^{\infty} A_k$ contracts to $\limsup A_n$, and $\bigcap_{k=n}^{\infty} A_k \subset A_n \subset \bigcup_{k=n}^{\infty} A_k$.
(e) Show that $(\liminf_n A_n)^c = \limsup_n A_n^c$, and $(\limsup_n A_n)^c = \liminf_n A_n{}^c$.

5. Find $\limsup_n A_n$ and $\liminf_n A_n$ if $\Omega = E^1$ and

$$A_n = \left[0, 1 - \frac{1}{n}\right] \quad \text{if } n \text{ is even}$$

$$= \left[-1, \frac{1}{n}\right] \quad \text{if } n \text{ is odd}$$

6. Let $\Omega = E^2$ and take $A_n =$ the interior of the circle with center at $((-1)^n/n, 0)$ and radius 1. Find $\limsup_n A_n$ and $\liminf_n A_n$.

7. Let $x_1, x_2, \ldots$ be a sequence of real numbers, and let $A_n = (-\infty, x_n)$. What is the connection between $\limsup_n x_n$ and $\limsup_n A_n$ (similarly for lim inf)?

8. (Second Borel-Cantelli lemma) If $A_1, A_2, \ldots$ are *independent* events and $\sum_{n=1}^{\infty} P(A_n) = \infty$, show that $P(\limsup_n A_n) = 1$. HINT: Show that $P(\limsup_n A_n) = \lim_{n\to\infty} \lim_{m\to\infty} P(\bigcup_{k=n}^{m} A_k)$, and consider $(\bigcup_{k=n}^{m} A_k)^c$; use the fact that $e^{-x} \geq 1 - x$.

9. Let $R_1, R_2, \ldots$ be a sequence of independent random variables, all defined on the same probability space. Let c be any real number. Show that $R_n \xrightarrow{\text{a.s.}} c$ if and only if for every $\varepsilon > 0$, $\sum_{n=1}^{\infty} P\{|R_n - c| \geq \varepsilon\} < \infty$.

10. Let $R_1, R_2, \ldots$ be a sequence of independent random variables, and let c be any real number. Show that either $R_n \xrightarrow{\text{a.s.}} c$ or R_n "diverges almost surely" from c; that is, $P\{x: \lim_{n\to\infty} R_n(x) = c\} = 0$. Thus, for example, it is impossible that $P\{R_n \to c\} = 1/3$.

11. Let $R_1, R_2, \ldots$ be independent random variables, with $R_n = 1$ with probability p_n, $R_n = 0$ with probability $1 - p_n$.

(a) What conditions on the p_n are equivalent to the statement that $R_n \xrightarrow{P} 0$?

(b) What conditions on the p_n are equivalent to the statement that $R_n \xrightarrow{\text{a.s.}} 0$?

12. Let $R_1, R_2, \ldots$ be independent random variables, with $E(R_n) \equiv 0$, $\operatorname{Var} R_n = \sigma_n^2 \leq M/n$, where M is some fixed positive constant. Show that $(R_1 + \cdots + R_n)/n \xrightarrow{\text{a.s.}} 0$.

13. Give an example of a particular sequence of random variables $R_1, R_2, \ldots$ and a random variable R such that $0 < P\{R_n \to R\} < 1$.