STOCHASTIC PROCESSES

SHELDON M. ROSS

STOCHASTIC PROCESSES

SHELDON M. ROSS
University of California, Berkeley

JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto • Singapore

Library of Congress Cataloging in Publication Data:

Ross, Sheldon M.
Stochastic processes.

(Wiley series in probability and mathematical
statistics. Probability and mathematical statistics,
ISSN 0271-6232)
Includes bibliographical references and index.
1. Stochastic processes. I. Title. II. Series.
QA274.R65 1982 519.2 82-8619
ISBN 0-471-09942-2 AACR2

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

PREFACE

This text is a nonmeasure theoretic introduction to stochastic processes, and as such assumes a knowledge of calculus and elementary probability. In it we attempt to present some of the theory of stochastic processes, to indicate its diverse range of applications, and also to give the student some probabilistic intuition and insight in thinking about problems. We have attempted, wherever possible to view processes from a probabilistic instead of an analytic point of view. This attempt, for instance, has led us to study most processes from a sample path point of view.

I would like to thank Mark Brown, Cyrus Derman, Shun-Chen Niu, Michael Pinedo, and Zvi Schechner for their helpful comments.

SHELDON M. ROSS

CONTENTS

CHAPTER 1

Preliminaries

1.1 PROBABILITY

A basic notion in probability theory is *random experiment:* an experiment whose outcome cannot be determined in advance. The set of all possible outcomes of an experiment is called the *sample space* of that experiment, and we denote it by S.

An *event* is a subset of a sample space, and is said to occur if the outcome of the experiment is an element of that subset. We shall suppose that for each event E of the sample space S a number $P(E)$ is defined and satisfies the following three axioms*:

Axiom (1) $0 \leq P(E) \leq 1$.
Axiom (2) $P(S) = 1$.
Axiom (3) For any sequence of events $E_1, E_2, \ldots$ that are mutually exclusive, that is, events for which $E_iE_j = \phi$ when $i \neq j$ (where ϕ is the null set),

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

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

Some simple consequences of axioms (1), (2), and (3) are:

1.1.1. If $E \subset F$, then $P(E) \leq P(F)$.
1.1.2. $P(E^c) = 1 - P(E)$ where E^c is the complement of E.
1.1.3. $P(\bigcup_1^n E_i) = \sum_1^n P(E_i)$ when the E_i are mutually exclusive.
1.1.4. $P(\bigcup_1^\infty E_i) \leq \sum_1^\infty P(E_i)$.

The inequality (1.1.4) is known as *Boole's inequality.*

An important property of the probability function P is that it is continuous. To make this more precise, we need the concept of a limiting event, which we define as follows: A sequence of events $\{E_n, n \geq 1\}$ is said to be an *increasing* sequence if $E_n \subset E_{n+1}, n \geq 1$ and is said to be *decreasing* if $E_n \supset E_{n+1}, n \geq 1$.

* Actually $P(E)$ will only be defined for the so-called measureable events of S. But this restriction need not concern us.

If $\{E_n, n \geq 1\}$ is an increasing sequence of events, then we define a new event, denoted by $\lim_{n\to\infty} E_n$ by

$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i \qquad \text{when } E_n \subset E_{n+1}, \quad n \geq 1.$$

Similarly if $\{E_n, n \geq 1\}$ is a decreasing sequence, then defined $\lim_{n\to\infty} E_n$ by

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i, \qquad \text{when } E_n \supset E_{n+1}, \quad n \geq 1.$$

We may now state the following:

PROPOSITION 1.1.1

If $\{E_n, n \geq 1\}$ is either an increasing or decreasing sequence of events, then

$$\lim_{n\to\infty} P(E_n) = P\left(\lim_{n\to\infty} E_n\right).$$

Proof Suppose, first, that $\{E_n, n \geq 1\}$ is an increasing sequence, and define events F_n, $n \geq 1$ by

$$F_1 = E_1,$$

$$F_n = E_n\left(\bigcup_{1}^{n-1} E_i\right)^c = E_n E_{n-1}^c, \qquad n > 1.$$

That is, F_n consists of those points in E_n that are not in any of the earlier E_i, $i < n$. It is easy to verify that the F_n are mutually exclusive events such that

$$\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i \quad \text{and} \quad \bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i \qquad \text{for all } n \geq 1.$$

Thus

$$\begin{aligned} P\left(\bigcup_{1}^{\infty} E_i\right) &= P\left(\bigcup_{i}^{\infty} F_i\right) \\ &= \sum_{1}^{\infty} P(F_i) \qquad \text{(by Axiom 3)} \\ &= \lim_{n\to\infty} \sum_{1}^{n} P(F_i) \\ &= \lim_{n\to\infty} P\left(\bigcup_{1}^{n} F_i\right) \\ &= \lim_{n\to\infty} P\left(\bigcup_{1}^{n} E_i\right) \\ &= \lim_{n\to\infty} P(E_n), \end{aligned}$$

which proves the result when $\{E_n, n \geq 1\}$ is increasing.

If $\{E_n, n \geq 1\}$ is a decreasing sequence, then $\{E_n^c, n \geq 1\}$ is an increasing sequence; hence,

$$P\left(\bigcup_1^\infty E_n^c\right) = \lim_{n\to\infty} P(E_n^c).$$

But, as $\bigcup_1^\infty E_n^c = (\bigcap_1^\infty E_n)^c$, we see that

$$1 - P\left(\bigcap_1^\infty E_n\right) = \lim_{n\to\infty} [1 - P(E_n)],$$

or, equivalently,

$$P\left(\bigcap_1^\infty E_n\right) = \lim_{n\to\infty} P(E_n),$$

which proves the result.

Example 1.1(a). Consider a population consisting of individuals able to produce offspring of the same kind. The number of individuals initially present, denoted by X_0, is called the size of the zeroth generation. All offspring of the zeroth generation constitute the first generation and their number is denoted by X_1. In general, let X_n denote the size of the nth generation.

Since $X_n = 0$ implies that $X_{n+1} = 0$, it follows that $P\{X_n = 0\}$ is increasing and thus $\lim_{n\to\infty} P\{X_n = 0\}$ exists. What does it represent? To answer this use Proposition 1.1.1 as follows:

$$\begin{aligned}\lim_{n\to\infty} P\{X_n = 0\} &= P\left\{\lim_{n\to\infty} \{X_n = 0\}\right\}\\ &= P\left\{\bigcup_n \{X_n = 0\}\right\}\\ &= P\{\text{the population ever dies out}\}.\end{aligned}$$

That is, the limiting probability that the nth generation is void of individuals is equal to the probability of eventual extinction of the population.

Proposition 1.1.1 can also be used to prove the Borel–Cantelli lemma.

PROPOSITION 1.1.2. The Borel–Cantelli Lemma

Let $E_1, E_2, \ldots$ denote a sequence of events. If

$$\sum_{i=1}^\infty P(E_i) < \infty,$$

then

$$P\{\text{an infinite number of the } E_i \text{ occur}\} = 0.$$

Proof The event that an infinite number of the E_i occur, called the $\limsup_{i\to\infty} E_i$, can be expressed as

$$\limsup_{i\to\infty} E_i = \bigcap_{n=1}^\infty \bigcup_{i=n}^\infty E_i.$$

This follows since if an infinite number of the E_i occur, then $\bigcup_{i=n}^{\infty} E_i$ occurs for each n and thus $\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} E_i$ occurs. On the other hand, if $\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} E_i$ occurs, then $\bigcup_{i=n}^{\infty} E_i$ occurs for each n, and thus for each n at least one of the E_i occurs where $i \geq n$; and, hence, an infinite number of the E_i occur.

As $\bigcup_{i=n}^{\infty} E_i$, $n \geq 1$, is a decreasing sequence of events, it follows from Proposition 1.1.1 that

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} E_i\right) &= P\left(\lim_{n\to\infty} \bigcup_{i=n}^{\infty} E_i\right) \\ &= \lim_{n\to\infty} P\left(\bigcup_{i=n}^{\infty} E_i\right) \\ &\leq \lim_{n\to\infty} \sum_{i=n}^{\infty} P(E_i) \\ &= 0, \end{aligned}$$

and the result is proven.

Example 1.1(b). Let $X_1, X_2, \ldots$ be such that

$$P\{X_n = 0\} = 1/n^2 = 1 - P\{X_n = 1\}, \qquad n \geq 1.$$

If we let $E_n = \{X_n = 0\}$, then, as $\sum_1^{\infty} P(E_n) < \infty$, it follows from the Borel–Cantelli lemma that the probability that X_n equals 0 for an infinite number of n is equal to 0. Hence, for all n sufficiently large, X_n must equal 1, and so we may conclude that, with probability 1,

$$\lim_{n\to\infty} X_n = 1.$$

For a converse to the Borel–Cantelli lemma, independence is required.

PROPOSITION 1.1.3. Converse to the Borel–Cantelli Lemma

If $E_1, E_2, \ldots$ are independent events such that

$$\sum_{n=1}^{\infty} P(E_n) = \infty,$$

then

$$P\{\text{an infinite number of the } E_n \text{ occur}\} = 1.$$

Proof

$$\begin{aligned} P\{\text{an infinite number of the } E_n \text{ occur}\} &= P\left\{\lim_{n\to\infty} \bigcup_{i=n}^{\infty} E_i\right\} \\ &= \lim_{n\to\infty} P\left(\bigcup_{i=n}^{\infty} E_i\right) \\ &= \lim_{n\to\infty} \left[1 - P\left(\bigcap_{i=n}^{\infty} E_i^c\right)\right]. \end{aligned}$$

Now,

$$P\left(\bigcap_{i=n}^{\infty} E_i^c\right) = \prod_{i=n}^{\infty} P(E_i^c) \qquad \text{(by independence)}$$

$$= \prod_{i=n}^{\infty} (1 - P(E_i))$$

$$\leq \prod_{i=n}^{\infty} e^{-P(E_i)} \qquad \text{(by the inequality } 1 - x \leq e^{-x}\text{)}$$

$$= \exp\left(-\sum_{n}^{\infty} P(E_i)\right)$$

$$= 0 \quad \text{since } \sum_{i=n}^{\infty} P(E_i) = \infty \text{ for all } n.$$

Hence the result follows.

Example 1.1(c). Let $X_1, X_2, \ldots$ be independent and such that

$$P\{X_n = 0\} = 1/n = 1 - P\{X_n = 1\}, \qquad n \geq 1.$$

If we let $E_n = \{X_n = 0\}$, then as $\sum_{n=1}^{\infty} P(E_n) = \infty$ it follows from Proposition 1.1.3 that E_n occurs infinitely often. Also, as $\sum_{n=1}^{\infty} P(E_n^c) = \infty$ it also follows that E_n^c also occurs infinitely often. Hence, with probability 1, X_n will equal 0 infinitely often and will also equal 1 infinitely often. Hence, with probability 1, X_n will not approach a limiting value as $n \to \infty$.

1.2 RANDOM VARIABLES

Consider a random experiment having sample space S. A *random variable* X is a function that assigns a real value to each outcome in S. For any set of real numbers A, the probability that X will assume a value that is contained in the set A is equal to the probability that the outcome of the experiment is contained in $X^{-1}(A)$. That is,

$$P\{X \in A\} = P(X^{-1}(A)),$$

where $X^{-1}(A)$ is the event consisting of all points $s \in S$ such that $X(s) \in A$.

The *distribution function* F of the random variable X is defined for any real number x by

$$F(x) = P\{X \leq x\} = P\{X \in (-\infty, x)\}.$$

We shall denote $1 - F(x)$ by $\bar{F}(x)$, and so

$$\bar{F}(x) = P\{X > x\}.$$

A random variable X is said to be *discrete* if its set of possible values is countable. For discrete random variables,

$$F(x) = \sum_{y \leq x} P\{X = y\}.$$

A random variable is called *continuous* if there exists a function $f(x)$, called the *probability density function*, such that

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

for every set B. Since $F(x) = \int_{-\infty}^{x} f(x)\,dx$, it follows that

$$f(x) = \frac{d}{dx} F(x).$$

The *joint distribution function* F of two random variables X and Y is defined by

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

The distribution functions of X and Y,

$$F_X(x) = P\{X \le x\} \quad \text{and} \quad F_Y(x) = P\{Y \le x\},$$

can be obtained from $F(x, y)$ by making use of the continuity property of the probability operator. Specifically, let y_n, $n \ge 1$, denote an increasing sequence converging to ∞. Then as the events $\{X \le x, Y \le y_n\}$, $n \ge 1$, are increasing and

$$\lim_{n \to \infty} \{X \le x, Y \le y_n\} = \bigcup_{n=1}^{\infty} \{X \le x, Y \le y_n\} = \{X \le x\},$$

it follows from the continuity property that

$$\lim_{n \to \infty} P\{X \le x, Y \le y_n\} = P\{X \le x\},$$

or, equivalently,

$$F_X(x) = \lim_{y \to \infty} F(x, y).$$

Similarly,

$$F_Y(y) = \lim_{x \to \infty} F(x, y).$$

The random variables X and Y are said to be *independent* if

$$F(x, y) = F_X(x) F_Y(y)$$

for all x and y.

The random variables X and Y are said to be *jointly continuous* if there exists a function $f(x, y)$, called the *joint probability density function*, such that

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

for all sets A and B.

The joint distribution of any collection $X_1, X_2, \ldots, X_n$ of random variables is defined by

$$F(x_1, \ldots, x_n) = P\{X_1 \le x_1, \ldots, X_n \le x_n\}.$$

Furthermore, the n random variables are said to be independent if

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n),$$

where

$$F_{X_i}(x_i) = \lim_{\substack{x_j \to \infty \\ j \neq i}} F(x_1, \ldots, x_n).$$

1.3 EXPECTED VALUE

The *expectation* or *mean* of the random variable X, denoted by $E[X]$, is defined by

(1.3.1)
$$E[X] = \int_{-\infty}^{\infty} x\,dF(x) = \begin{cases} \int_{-\infty}^{\infty} x f(x)\,dx & \text{if } X \text{ is continuous} \\ \sum_x xP\{X = x\} & \text{if } X \text{ is discrete} \end{cases}$$

provided the above integral exists.

Equation (1.3.1) also defines the expectation of any function of X, say $h(X)$. Since $h(X)$ is itself a random variable, it follows from (1.3.1) that

$$E[h(X)] = \int_{-\infty}^{\infty} x\,dF_h(x),$$

where F_h is the distribution function of $h(X)$. However, it can be shown that this is identical to $\int_{-\infty}^{\infty} h(x)\,dF(x)$. That is,

(1.3.2)
$$E[h(X)] = \int_{-\infty}^{\infty} h(x)\,dF(x).$$

The above equation is sometimes known as the *law of the unconscious statistician* [since statisticians have been accused of using the identity (1.3.2) without realizing that it is not a definition].

The variance of the random variable X is defined by

$$\begin{aligned} \operatorname{Var} X &= E[(X - E[X])^2] \\ &= E[X^2] - E^2[X]. \end{aligned}$$

Two jointly distributed random variables X and Y are said to be uncorrelated if their covariance, defined by

$$\begin{aligned} \operatorname{Cov}(X, Y) &= E[(X - EX)(Y - EY)] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

is zero. It follows that independent random variables are uncorrelated. However, the converse need not be true. (The reader should think of an example.)

An important property of expectations is that the expectation of a sum of random variables is equal to the sum of the expectations.

(1.3.3)
$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i].$$

The corresponding property for variances is that

$$\text{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum\sum_{i<j} \text{Cov}(X_i, X_j). \tag{1.3.4}$$

Example 1.3(a). The Matching Problem. At a party n people put their hats in the center of a room where the hats are mixed together. Each person then randomly selects one. We are interested in the mean and variance of X—the number that select their own hat.

To solve we use the representation

$$X = X_1 + X_2 + \cdots + X_n,$$

where

$$X_i = \begin{cases} 1 & \text{if the } i\text{th person selects his own hat} \\ 0 & \text{otherwise.} \end{cases}$$

Now, as the ith person is equally likely to select any of the n hats, it follows that $P\{X_i = 1\} = 1/n$, and so

$$E[X_i] = 1/n,$$

$$\text{Var}(X_i) = \frac{1}{n}\left(1 - \frac{1}{n}\right) = \frac{n-1}{n^2}.$$

Also

$$\text{Cov}(X_i, X_j) = E[X_iX_j] - E[X_i]E[X_j].$$

Now,

$$X_iX_j = \begin{cases} 1 & \text{if the } i\text{th and } j\text{th men both select their own hats} \\ 0 & \text{otherwise,} \end{cases}$$

and thus

$$\begin{aligned} E[X_iX_j] &= P\{X_i = 1, X_j = 1\} \\ &= P\{X_i = 1\}P\{X_j = 1 \mid X_i = 1\} \\ &= \frac{1}{n}\frac{1}{n-1}. \end{aligned}$$

Hence,

$$\text{Cov}(X_i, X_j) = \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2 = \frac{1}{n^2(n-1)}.$$

Therefore, from (1.3.3) and (1.3.4),

$$E[X] = 1$$

and

$$\begin{aligned} \text{Var}(X) &= \frac{n-1}{n} + 2\binom{n}{2}\frac{1}{n^2(n-1)} \\ &= 1. \end{aligned}$$

Thus both the mean and variance of the number of matches are equal to 1. (See Example 1.5(e) for an explanation as to why these results are not surprising.)

Example 1.3(b). Some Probability Identities. Let $A_1, A_2, \ldots, A_n$ denote events and define the indicator variables $I_j, j = 1, \ldots, n$ by

$$I_j = \begin{cases} 1 & \text{if } A_j \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

Letting

$$N = \sum_{j=1}^{n} I_j,$$

then N denotes the number of the A_j, $1 \le j \le n$, that occur. A useful identity can be obtained by noting that

$$(1.3.5) \qquad (1-1)^N = \begin{cases} 1 & \text{if } N = 0 \\ 0 & \text{if } N > 0. \end{cases}$$

But by the binomial theorem

$$(1.3.6) \qquad (1-1)^N = \sum_{i=0}^{N} \binom{N}{i} (-1)^i$$

$$= \sum_{i=0}^{n} \binom{N}{i} (-1)^i \qquad \text{since } \binom{m}{i} = 0 \text{ when } i > m.$$

Hence, if we let

$$I = \begin{cases} 1 & \text{if } N > 0 \\ 0 & \text{if } N = 0, \end{cases}$$

then (1.3.5) and (1.3.6) yield

$$1 - I = \sum_{i=0}^{n} \binom{N}{i} (-1)^i$$

or

$$(1.3.7) \qquad I = \sum_{i=1}^{n} \binom{N}{i} (-1)^{i+1}.$$

Taking expectations of both sides of (1.3.7) yields

$$(1.3.8) \qquad E[I] = E[N] - E\left[\binom{N}{2}\right] + \cdots + (-1)^{n+1} E\left[\binom{N}{n}\right].$$

However,

$$E[I] = P\{N > 0\}$$

$$= P\{\text{at least one of the } A_i \text{ occurs}\}$$

$$= P\left(\bigcup_1^n A_i\right)$$

and

$$E[N] = E\left[\sum_{j=1}^{n} I_j\right] = \sum_{j=1}^{n} P(A_j),$$

$$E\left[\binom{N}{2}\right] = E[\text{number of pairs of the } A_j \text{ that occur}]$$

$$= E\left[\sum\sum_{i<j} I_i I_j\right]$$

$$= \sum\sum_{i<j} E[I_i I_j]$$

$$= \sum\sum_{i<j} P(A_i A_j),$$