Topics in Stochastic Processes

ROBERT B. ASH
MELVIN F. GARDNER

Topics in Stochastic Processes

ROBERT B. ASH

Department of Mathematics
University of Illinois
Urbana, Illinois

MELVIN F. GARDNER

Department of Mathematics
University of Toronto
Toronto, Ontario, Canada

ACADEMIC PRESS New York San Francisco London 1975

A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

Library of Congress Cataloging in Publication Data

Ash, Robert B
Topics in stochastic processes.

(Probability and mathematical statistics series
Bibliography: p.
Includes index.
1. Stochastic processes. I. Gardner, Melvin F., joint author. II. Title.
QA274.A79 519.2 74-17991
ISBN 0-12-065270-6
AMS (MOS) 1970 Subject Classification: 60-62

PRINTED IN THE UNITED STATES OF AMERICA

79 80 81 82 9 8 7 6 5 4 3 2

Preface

This book contains selected topics in stochastic processes that we believe can be studied profitably by a reader familiar with basic measure-theoretic probability. The background is given in "Real Analysis and Probability" by Robert B. Ash, Academic Press, 1972. A student who has learned this material from other sources will be in good shape if he feels reasonably comfortable with infinite sequences of random variables. In particular, a reader who has studied versions of the strong law of large numbers and the central limit theorem, as well as basic properties of martingale sequences, should find our presentation accessible.

We should comment on our choice of topics. In using the tools of measure-theoretic probability, one is unavoidably operating at a high level of abstraction. Within this limitation, we have tried to emphasize processes that have a definite physical interpretation and for which explicit numerical results can be obtained, if desired. Thus we begin (Chapters 1 and 2) with L^2 stochastic processes and prediction theory. Once the underlying mathematical foundation has been built, results which have been used for many years by engineers and physicists are obtained. The main result of Chapter 3, the ergodic theorem, may be regarded as a version of the strong law of large numbers for stationary stochastic processes. We describe several interesting applications to real analysis, Markov chains, and information theory.

In Chapter 4 we discuss the sample function behavior of continuous parameter processes. General properties of martingales and Markov processes are given, and one-dimensional Brownian motion is analyzed in detail. The purpose is to illustrate those concepts and constructions that are basic in any discussion of continuous parameter processes, and to open the gate to allow the reader to proceed to more advanced material on Markov processes and potential theory. In Chapter 5 we use the theory of continuous parameter processes to develop the Itô stochastic integral and to discuss the solution of stochastic differential equations. The results are of current interest in communication and control theory.

The text has essentially three independent units: Chapters 1 and 2; Chapter 3; and Chapters 4 and 5. The system of notation is standard; for example, 2.3.1 means Chapter 2, Section 3, Part 1. A reference to "Real Analysis and Probability" is denoted by RAP.

Problems are given at the end of each section. Fairly detailed solutions are given to many problems.

We are indebted to Mary Ellen Bock and Ed Perkins for reading the manuscript and offering many helpful suggestions.

Once again we thank Mrs. Dee Keel for her superb typing, and the staff of Academic Press for their constant support and encouragement.

Contents

Chapter 1

L^2 Stochastic Processes

1.1 Introduction

We shall begin our work in stochastic processes by considering a down-to-earth class of such processes, those whose random variables have finite second moments. The objective of this chapter is to develop some intuition in handling stochastic processes, and to prepare for the study of spectral theory and prediction in Chapter 2.

First we must recall some notation from probability and measure theory.

1.1.1 Terminology

A *measurable space* $(S, \mathscr{S})$ is a set S with a σ-field $\mathscr{S}$ of subsets of S. A *probability space* $(\Omega, \mathscr{F}, P)$ is a measure space $(\Omega, \mathscr{F})$ with a probability measure P on $\mathscr{F}$. If X is a function from Ω to S, X is said to be *measurable* (notation $X: (\Omega, \mathscr{F}) \to (S, \mathscr{S})$) iff $X^{-1}(A) \in \mathscr{F}$ for each $A \in \mathscr{S}$. If S is the set R of reals and $\mathscr{S} = \mathscr{B}(R)$, the class of Borel sets of R, X is called a *random variable* or, for emphasis, a *real random variable*; if $S = \bar{R}$ (the extended reals) and $\mathscr{S} = \mathscr{B}(\bar{R})$, X is said to be an *extended random variable*; if $S = C$ (the complex numbers) and $\mathscr{S} = \mathscr{B}(C)$; X is called a *complex random variable*. More generally, if S is the set R^n of n-tuples of real numbers, and $\mathscr{S} = \mathscr{B}(R^n)$, X is called an *n-dimensional random vector*; X may be regarded as an n-tuple $(X_1, \ldots, X_n)$ of random variables (RAP, p. 212). If $S = R^\infty$, the set of all sequences of real numbers, and $\mathscr{S} = \mathscr{B}(R^\infty)$, X is said to be a *random sequence*; X may be regarded as a sequence $(X_1, X_2, \ldots)$ of random variables (RAP, 5.11.3, p. 233).

We now give the general definition of a stochastic process, and the probability student will see that he has already encountered many examples of such processes.

1.1.2 Definitions and Comments

Let $(\Omega, \mathscr{F}, P)$ be a probability space, $(S, \mathscr{S})$ an arbitrary measurable space, and T an arbitrary set. A *stochastic process* on $(\Omega, \mathscr{F}, P)$ with *state space* $(S, \mathscr{S})$ and *index set* T is a family of measurable functions X_t: $(\Omega, \mathscr{F}) \to (S, \mathscr{S})$, $t \in T$.

Note that if T is the set of positive integers and $S = R$, $\mathscr{S} = \mathscr{B}(R)$, the stochastic process $\{X_t, t \in T\}$ is a sequence of random variables. Similarly, we can obtain n-dimensional random vectors $(T = \{1, \ldots, n\}, S = R$, $\mathscr{S} = \mathscr{B}(R))$ and even single random variables $(n = 1)$.

A synonym for stochastic process is *random function*; let us try to explain this terminology. Let S^T be the collection of all functions from T to S, and let $\mathscr{S}^T$ be the *product σ-field* on S^T. (Recall that $\mathscr{S}^T$ is the smallest σ-field containing all measurable rectangles

$$\{\omega \in S^T : \omega(t_1) \in B_1, \ \ldots, \omega(t_n) \in B_n\},$$

$t_1, \ldots, t_n \in T$, $B_1, \ldots, B_n \in \mathscr{S}$, $n = 1, 2, \ldots$; see RAP, 4.4.1, p. 189.)

Now suppose that for each $t \in T$ we have a function X_t: $\Omega \to S$; define $X : \Omega \to S^T$ by

$$X(\omega) = (X_t(\omega), t \in T).$$

Then X is a measurable map of $(\Omega, \mathscr{F})$ into $(S^T, \mathscr{S}^T)$ iff each X_t is a measurable map of $(\Omega, \mathscr{F})$ into $(S, \mathscr{S})$ (see Problem 1).

Thus a stochastic process is a measurable mapping X from Ω into the function space S^T. Intuitively, the performance of the experiment produces a sample point ω, and this in turn determines a collection of elements $X(t, \omega) = X_t(\omega)$, $t \in T$. In other words, the outcome of the experiment determines a function from T to S, called the *sample function* corresponding to the point ω. If T is an interval of reals, it is often helpful to visualize t as a time parameter, and to think of a stochastic process as a random time function.

In this chapter we shall be concerned with L^2 processes, defined as follows. An *L^2 stochastic process* is a family of real or complex random variables X_t, $t \in T$, such that $\|X_t\|^2 = E(|X_t|^2) < \infty$ for all $t \in T$. Thus the state space is R or C, and we have a second moment restriction.

1.1.3 Examples

(a) Let θ be a random variable, uniformly distributed between 0 and 2π. (We may take $\Omega = [0, 2\pi)$, $\mathscr{F} = \mathscr{B}[0, 2\pi)$, $P(B) = \int_B (1/2\pi)\,dx$, $\theta(\omega) = \omega$.)

We define a stochastic process by $X_t = \sin(at + \theta)$ where a is fixed and t ranges over all real numbers. Explicitly, $X_t(\omega) = \sin(at + \theta(\omega))$, $\omega \in \Omega$. Thus the process represents a sine wave with a random phase angle (see Figure 1.1).

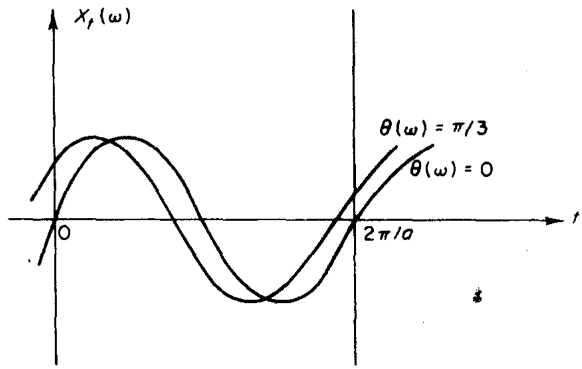

FIGURE 1.1

(b) Let V and W be independent random variables with distribution functions F and G (take $\Omega = R^2$, $\mathscr{F} = \mathscr{B}(R^2)$, $P(B) = \iint_B dF(x)\,dG(y)$, $V(x, y) = x$, $W(x, y) = y$). For t real, let $X_t = 0$ for $t < V$, $X_t = W$ for $t \geq V$, that is, $X_t(\omega) = 0$ for $t < V(\omega)$, $X_t(\omega) = W$ for all $t \geq V(\omega)$. The process represents a step function with random starting time and random amplitude (see Figure 1.2).

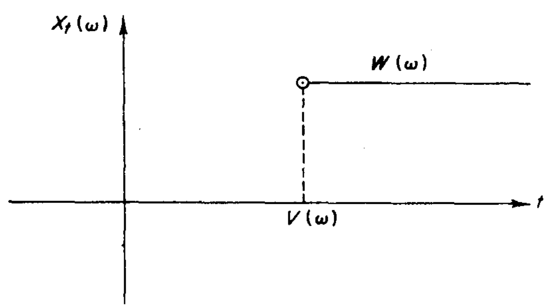

FIGURE 1.2

We now describe a basic approach to the construction of stochastic processes.

1.1.4 The Kolmogorov Construction

If $\{X_t, t \in T\}$ is a stochastic process, we may compute the distribution of $(X_{t_1}, \ldots, X_{t_n})$ for each $t_1, \ldots, t_n \in T$, $t_1 < \cdots < t_n$. (If T is not a subset of the reals, a fixed total ordering is assumed on T. This avoids the problem of having to deal with all permutations of $t_1, \ldots, t_n$.) For example, in 1.1.3(b) we have

$$P\{X_2 > 1, 2 \le X_5 \le 4, X_{27} > 3\} = P\{V \le 2, 3 < W \le 4\}$$
$$= F(2)(G(4) - G(3)).$$

In fact if we specify in a consistent manner the distribution of $(X_{t_1}, \ldots, X_{t_n})$ for all finite subsets $\{t_1, \ldots, t_n\}$ of T, it is possible to construct a stochastic process with the given finite-dimensional distributions. This is a consequence of the Kolmogorov extension theorem (RAP, 4.4.3, p. 191). To see how the theorem is applied, let $\Omega = S^T$, $\mathscr{F} = \mathscr{S}^T$; we assume throughout that S is a complete, separable metric space and $\mathscr{S} = \mathscr{B}(S)$, the class of Borel sets of S. For each finite subset $v = \{t_1, \ldots, t_n\}$ of T, suppose that we specify a probability measure P_v on $\mathscr{S}^n$; $P_v(B)$ is to represent $P\{\omega \in \Omega : (X_{t_1}(\omega), \ldots, X_{t_n}(\omega)) \in B\}$, which will equal $P\{\omega \in \Omega : (\omega(t_1), \ldots, \omega(t_n)) \in B\}$ if we take $X_t(\omega) = \omega(t)$, $t \in T$.

The hypothesis of the Kolmogorov extension theorem requires that the P_v be consistent, that is, if $u = \{\tau_1, \ldots, \tau_k\} \subset v = \{t_1, \ldots, t_n\}$, the projection $\pi_u(P_v)$ (RAP, p. 190) must coincide with P_v. Now

$$[\pi_u(P_v)](B) = P_v\{y = (y(t_1), \ldots, y(t_n)) \in S^n : (y(\tau_1), \ldots, y(\tau_k)) \in B\}.$$

In terms of finite-dimensional distributions, $\pi_u(P_v)$ gives the distribution of $(X_{\tau_1}, \ldots, X_{\tau_k})$ *as calculated from the distribution of the larger family* $(X_{t_1}, \ldots, X_{t_n})$, whereas P_u gives the distribution of $(X_{\tau_1}, \ldots, X_{\tau_k})$ as originally specified. Thus the consistency requirement is quite natural. For example, if we want X_1, X_2, and X_3 to be independent, normally distributed random variables, we cannot at the same time demand that X_1 be uniformly distributed; it must be normal.

Now let us recall the precise statement of the Kolmogorov extension theorem:

Assume that for each finite nonempty subset v of T we are given a probability measure P_v on $\mathscr{S}^n$, where n is the number of elements in v; assume also that the P_v are consistent. Then there is a unique probability measure P on $\mathscr{S}^T$ such that $\pi_v(P) = P_v$ for all v, that is, if $v = \{t_1, \ldots, t_n\}$, $B \in \mathscr{S}^n$, we have

$$P\{\omega : (\omega(t_1), \ldots, \omega(t_n)) \in B\} = P_v(B).$$

Thus if we set $X_t(\omega) = \omega(t)$, $t \in T$ (or equivalently, if we take X to be the identity map on S^T), we have produced a stochastic process with the given finite-dimensional distributions.

Now suppose that $\{Y_t, t \in T\}$ is another process (on a possibly different probability space $(\Omega', \mathscr{F}', P')$) with the same finite-dimensional distributions. Define the mapping $Y: \Omega' \to S^T$ by $Y(\omega) = (Y_t(\omega), t \in T)$; then (Problem 1) Y is a measurable function from $(\Omega', \mathscr{F}')$ to $(S^T, \mathscr{S}^T)$. If $B \in \mathscr{S}^T$, we assert that $P_X(B) = P_Y(B)$; this holds when B is a measurable cylinder since the two processes have the same finite-dimensional distributions, and the general result follows from the Carathéodory extension theorem. The statement that $P_X = P_Y$ is the precise version of the intuitive statement that the finite-dimensional distributions determine the distribution of the entire process.

The above discussion applies with only minor notational changes if the fixed state space $(S, \mathscr{S})$ is replaced by a family of spaces $(S_t, \mathscr{S}_t)$ where each S_t is a complete separable metric space and $\mathscr{S}_t$ is the class of Borel sets.

We shall often use the Kolmogorov extension theorem to construct stochastic processes in situations where the initial data consist of the finite-dimensional distributions. In particular, if we specify that $X_{t_1}, \ldots, X_{t_n}$ be independent whenever $t_1 < \cdots < t_n$, the consistency requirement is satisfied, so that it is possible to construct a family of independent random variables X_t with arbitrary distribution functions F_t, $t \in T$. In some cases, more complicated stochastic processes are constructed using the independent random variables X_t; the following example illustrates this idea.

1.1.5 Example

Suppose that "customers" arrive at times T_1, $T_1 + T_2$, $\ldots$, $T_1 + \cdots + T_n$, $\ldots$, where the T_i are independent, strictly positive random variables, T_1 having distribution function F_1 and the T_i, $i > 1$, each having distribution function F.

Let $Y_n = T_1 + \cdots + T_n$ be the arrival time of the nth customer, and let $N(t)$ be the number of customers arriving in the interval $(0, t]$. Formally,

$$N(t) = k \quad \text{if} \quad Y_k \leq t < Y_{k+1}, \quad k = 0, 1, \ldots \quad (\text{define } Y_0 \equiv 0).$$

For any $t \geq 0$, let Z_t be the waiting time from t to the arrival of the next customer. We claim that

$$\{Z_t \leq x\} \overset{\text{a.e.}}{=} \bigcup_{n=1}^{\infty} \{t < Y_n \leq t + x < Y_{n+1}\}. \tag{1}$$

(We say that the sets A and B are equal a.e. iff $P(A \,\triangle\, B) = 0$.) For if $t < Y_n \leq t + x < Y_{n+1}$, the nth customer arrives in $(t, t + x]$, so $Z_t \leq x$. Conversely, if $Z_t \leq x$, then some customer arrives in $(t, t + x]$, and hence (a.e.) there is a last customer to arrive in this interval. (If not, then $\sum_{n=1}^{\infty} T_n < \infty$, and this has probability zero by the strong law of large numbers.) If customer n is the last to arrive in $(t, t + x]$, then $t < Y_n \le t + x < Y_{n+1}$.

We now find the distribution of Z_t. By (1),

$$P\{Z_t \le x\} = \sum_{n=1}^{\infty} P\{t < Y_n \le t + x < Y_{n+1}\}. \tag{2}$$

But

$$P\{t < Y_n \le t + x < Y_{n+1}\} = P\{t < Y_n \le t + x,\ Y_n + T_{n+1} > t + x\}$$

$$= \iint\limits_{\substack{t<u\le t+x \\ u+v>t+x}} dF_{Y_n}(u)\, dF_{T_{n+1}}(v)$$

$$= \int\limits_{t<u\le t+x} P\{T_{n+1} > t + x - u\}\, dF_{Y_n}(u),$$

as would be expected by a conditioning argument. Thus

$$F_{Z_t}(x) = \sum_{n=1}^{\infty} \int_{(t, t+x]} [1 - F(t + x - u)]\, dF_{Y_n}(u). \tag{3}$$

We now assume that the T_n, $n > 1$, have finite expectation m, and

$$F_1(x) = \frac{1}{m} \int_0^x [1 - F(y)]\, dy. \tag{4}$$

(This forces the arrival rate, that is, the average number of customers arriving per second, to be constant; see Problem 4. We make the assumption in order to obtain F_{Z_t} explicitly.)

By RAP (p. 280, Problem 2), $F_1(\infty) = 1$, and hence F_1 is the distribution function of a random variable.

Now if X is a random variable with distribution function G, the *generalized characteristic function* of X (or of G) is defined by

$$M(s) = \int_R e^{-sx}\, dG(x). \tag{5}$$

$M(s)$ is defined for those complex numbers s for which the integral is finite. If we set $s = -iu$, we obtain the "ordinary" characteristic function. The proof given in RAP (8.1.2, p. 322) shows that if $X_1, \ldots, X_n$ are independent random variables whose generalized characteristic functions $M_1, \ldots, M_n$ are defined at s, then the generalized characteristic function M of $X_1 + \cdots + X_n$ is defined at s, and

$$M(s) = \prod_{i=1}^{n} M_i(s). \tag{6}$$

Returning to the original problem, let M_1 and M be the generalized characteristic functions of T_1 and T_n, $n > 1$; we have, for Re $s > 0$,

$$M_1(s) = \int_0^\infty e^{-sx}\, dF_1(x) = \frac{1}{m}\int_0^\infty e^{-sx}(1 - F(x))\, dx. \tag{7}$$

We now assume that the T_n, $n > 1$, have a density f (this assumption can be avoided, but it simplifies the calculations). Note that by (4), T_1 has density $f_1(x) = m^{-1}(1 - F(x))$, $x \geq 0$. Then (7) becomes

$$\begin{aligned} M_1(s) &= \frac{1}{ms} - \frac{1}{m}\int_0^\infty e^{-sx}\int_0^x f(y)\, dy\, dx \\ &= \frac{1}{ms} - \frac{1}{m}\int_0^\infty f(y)\int_y^\infty e^{-sx}\, dx\, dy \\ &= \frac{1}{ms} - \frac{1}{ms}\int_0^\infty f(y)e^{-sy}\, dy. \end{aligned}$$

Thus

$$M_1(s) = \frac{1}{ms}(1 - M(s)),$$

or

$$\frac{1}{ms} = \frac{M_1(s)}{1 - M(s)} = M_1(s)\sum_{n=0}^{\infty}[M(s)]^n.$$

But the generalized characteristic function of Y_n is $M_1(s)[M(s)]^{n-1}$, hence

$$\frac{1}{ms} = \sum_{n=1}^{\infty} M_{Y_n}(s). \tag{8}$$

Since $T_1, \ldots, T_n$ have densities, so does Y_n. Thus

$$\sum_{n=1}^{\infty} M_{Y_n}(s) = \sum_{n=1}^{\infty}\int_0^\infty e^{-sx} f_{Y_n}(x)\, dx,$$

hence by (8),

$$\frac{1}{ms} = \int_0^\infty e^{-sx}\left[\sum_{n=1}^{\infty} f_{Y_n}(x)\right] dx. \tag{9}$$

But $(ms)^{-1} = \int_0^\infty e^{-sx}m^{-1}\,dx$, and it follows from the uniqueness theorem for Laplace transforms (Problem 7) that

$$\sum_{n=1}^{\infty} f_{Y_n}(x) = \frac{1}{m} \qquad \text{a.e. on} \quad [0, \infty) \quad \text{[Lebesgue measure].} \tag{10}$$

By (3),

$$\begin{aligned} F_{Z_t}(x) &= \frac{1}{m}\int_t^{t+x} [1 - F(t + x - u)]\,du \\ &= \frac{1}{m}\int_0^x [1 - F(y)]\,dy \qquad (\text{set } y = t + x - u) \\ &= F_1(x). \end{aligned}$$

Thus if we start counting at any time t, the waiting time to the next customer has exactly the same distribution as the initial waiting time starting at $t = 0$.

(If we do not assume that the T_n, $n > 1$, have a density, the above analysis may be carried through by replacing expressions of the form $f_{Y_n}(x)\,dx$ by $dF_{Y_n}(x)$. Equation (10) becomes $\sum_{n=1}^{\infty} F_{Y_n}(x) = x/m$ a.e.; this is proved using the uniqueness theorem for the Laplace transform of a distribution function (see, for example, Widder, 1941).)

In fact if $W_1, W_2, \ldots$ are the successive waiting times starting at t (so $W_1 = Z_t$), then $W_1, W_2, \ldots$ are independent, and W_i has the same distribution as T_i for all i. To see this, note that

$$\{W_1 \le x_1, \ldots, W_k \le x_k\} = \bigcup_{n=0}^{\infty} \{Y_n \le t < Y_{n+1} \le t + x_1, \\ T_{n+2} \le x_2, \ldots, T_{n+k} \le x_k\}. \tag{11}$$

For it is clear that the set on the right-hand side of (11) is a subset of the set on the left. Conversely, if $W_1 \le x_1, \ldots, W_k \le x_k$, then a customer arrives in $(t, t + x_1]$, hence there is a first customer in this interval, say customer $n + 1$. Then $Y_n \le t < Y_{n+1} \le t + x_1$, and also $W_i = T_{n+i}$, $i = 2, \ldots, k$, as desired. Therefore

$$\begin{aligned} P\{W_1 \le x_1, \ldots, W_k \le x_k\} &= \left[\sum_{n=0}^{\infty} P\{Y_n \le t < Y_{n+1} \le t + x_1\}\right] \prod_{i=2}^{k} F(x_i) \\ &= P\{Z_t \le x_1\} \prod_{i=2}^{k} F(x_i). \end{aligned}$$

Fix j and let $x_i \to \infty$, $i \neq j$, to conclude that the W_i are independent and W_i has the same distribution as T_i for all i. In particular, $N(t+h) - N(t)$ has the same distribution as $N(h)$.

The stochastic process $\{N(t), t \geq 0\}$ is sometimes called a *delayed renewal process*. (The word *delayed* refers to the fact that T_1 has a different distribution from the T_n, $n > 1$; if the distributions are the same, *delayed* is omitted.)

Physically, the T_i can be regarded as the lifetimes of a succession of products such as light bulbs. If $T_1 + \cdots + T_n = t$, then bulb n has burned out at time t, and the light must be renewed by placing bulb $n+1$ in position.

If F_1 and F are related as in (4), then $\{N(t), t \geq 0\}$ is called a *uniform* (delayed) renewal process. The most important example is the *Poisson process*, obtained by specifying that the T_n, $n > 1$, have the exponential density $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$ (where $\lambda > 0$). Then $m = 1/\lambda$, so by (4),

$$F_1(x) = \lambda \int_0^x [1 - (1 - e^{-\lambda y})]\, dy = 1 - e^{-\lambda x}, \qquad x \geq 0.$$

Therefore $F_1 = F$.

The reason for the use of the name *Poisson* is that for each t, $N(t)$ has a Poisson distribution with parameter λt, that is,

$$P\{N(t) = k\} = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \qquad k = 0, 1, \ldots.$$

To see this, observe that

$$P\{N(t) \leq k\} = P\{T_1 + \cdots + T_{k+1} > t\},$$

and $T_1 + \cdots + T_{k+1}$ has density $\lambda^{k+1} x^k e^{-\lambda x}/k!$, $x \geq 0$. (This is an exercise in basic probability theory; see Ash, 1970, p. 197 for details.) Thus

$$P\{N(t) \leq k\} = \int_t^\infty \frac{1}{k!} \lambda^{k+1} x^k e^{-\lambda x}\, dx.$$

Successive integration by parts (again see Ash, 1970, p. 197) yields

$$P\{N(t) \leq k\} = \sum_{i=0}^{k} e^{-\lambda t} \frac{(\lambda t)^i}{i!},$$

as desired.

Since $E[N(t)] = \lambda t$, $\lambda = 1/E(T_i)$ may be interpreted as the average number of customers per second. Furthermore, since $\operatorname{Var} N(t) = \lambda t$, $\{N(t), t \geq 0\}$ is an L^2 process.

A key property of the Poisson process is the following.

1.1.6 Theorem

Let $\{N(t),\ t \geq 0\}$ be a Poisson process. The process has *independent increments*, that is, if $0 < t_1 < \cdots < t_m$, then $N(t_1) - N(0)$, $N(t_2) - N(t_1)$, $\ldots$, $N(t_m) - N(t_{m-1})$ are independent; equivalently, if

$$0 \leq t_1 < t_2 \leq t_3 < t_4 \leq \cdots \leq t_{2n-1} < t_{2n},$$

then $N(t_2) - N(t_1)$, $N(t_4) - N(t_3)$, $\ldots$, $N(t_{2n}) - N(t_{2n-1})$ are independent.

PROOF We break the argument into several parts. Fix $t_1 > 0$, and let W_1, $W_2, \ldots$ be the successive waiting times starting from t_1, as in 1.1.5.

(a) $N(t_1)$ and W_1 are independent.
We do this by an induction argument. First note that

$$\begin{aligned} P\{N(t_1) = 0,\ W_1 \leq x_1\} &= P\{t_1 < T_1 \leq t_1 + x_1\} \\ &= e^{-\lambda t_1} - e^{-\lambda(t_1 + x_1)} = e^{-\lambda t_1}(1 - e^{-\lambda x_1}) \\ &= P\{N(t_1) = 0\}P\{W_1 \leq x_1\}. \end{aligned}$$

If $P\{N(t_1) = n - 1,\ W_1 \leq x_1\} = P\{N(t_1) = n - 1\}P\{W_1 \leq x_1\}$ for all $t_1 > 0$, $x_1 \geq 0$, then

$$\begin{aligned} P\{N(t_1) = n,\ W_1 \leq x_1\} &= P\{T_1 + \cdots + T_n \leq t_1 \\ &\qquad < T_1 + \cdots + T_{n+1} \leq t_1 + x_1\} \\ &= \int \cdots \int_A \lambda^{n+1} e^{-\lambda(y_1 + \cdots + y_{n+1})}\, dy_1 \cdots dy_{n+1} \end{aligned}$$

where

$$\begin{aligned} A = \{(y_1, \ldots, y_{n+1})&: 0 \leq y_1 \leq t_1,\ y_2, \ldots, y_{n+1} \geq 0, \\ &y_2 + \cdots + y_n \leq t_1 - y_1 < y_2 + \cdots + y_{n+1} \leq t_1 + x_1 - y_1\}. \end{aligned}$$

Thus

$$\begin{aligned} &P\{N(t_1) = n,\ W_1 \leq x_1\} \\ &\quad = \int_0^{t_1} \lambda e^{-\lambda y_1} P\{T_2 + \cdots + T_n \leq t_1 - y_1 \\ &\qquad\qquad < T_2 + \cdots + T_{n+1} \leq t_1 + x_1 - y_1\}\, dy_1 \\ &\quad = \int_0^{t_1} \lambda e^{-\lambda y_1} P\{N(t_1 - y_1) = n - 1,\ W_1^* \leq x_1\}\, dy_1 \end{aligned}$$

where W_1^* is the waiting time starting from $t_1 - y_1$. By the induction