John Lamperti

Stochastic Processes

A Survey of the Mathematical Theory

John Lamperti

Stochastic Processes

A Survey of the Mathematical Theory

Springer-Verlag
New York Heidelberg Berlin

John Lamperti
Department of Mathematics
Dartmouth College
Hanover, New Hampshire 03755

AMS Subject Classifications: 60-01, 60Gxx, 60Jxx

Library of Congress Cataloging in Publication Data

Lamperti, John.
Stochastic processes.

(Applied mathematical sciences ; v. 23)
Bibliography: p.
Includes index.
1. Stochastic processes. 2. Stationary processes. 3. Markov processes. I. Title.
II. Series.
QA1.A647 vol. 23 [QA274] 510'.8s [519.2]
77-24321

9 8 7 6 5 4 3 2 1

Printed in the United States of America.

ISBN 0-387-90275-9 Springer-Verlag New York Heidelberg Berlin

ISBN 3-540-90275-9 Springer-Verlag Berlin Heidelberg New York

PREFACE

This book is the result of lectures which I gave during the academic year 1972-73 to third-year students at Aarhus University in Denmark. The purpose of the book, as of the lectures, is to survey some of the main themes in the modern theory of stochastic processes.

In my previous book Probability: a survey of the mathematical theory I gave a short overview of "classical" probability mathematics, concentrating especially on sums of independent random variables. I did not discuss specific applications of the theory; I did strive for a spirit friendly to application by coming to grips as fast as I could with the major problems and techniques and by avoiding too high levels of abstraction and completeness. At the same time, I tried to make the proofs both rigorous and motivated and to show how certain results have evolved rather than just presenting them in polished final form. The same remarks apply to this book, at least as a statement of intentions, and it can serve as a sequel to the earlier one continuing the story in the same style and spirit.

The contents of the present book fall roughly into two parts. The first deals mostly with stationary processes, which provide the mathematics for describing phenomena in a steady state overall but subject to random fluctuations. Chapter 4 is the heart of this part. The simple geometry of the Wold decomposition is the starting point for discussing linear prediction, and the analysis is derived from it in a direct and natural way. Basic results such as criteria for singularity or regularity, the factorization of the spectral density and the error of optimum prediction are thus obtained with a minimum of heavy analytical machinery, while the need for those tools which are used becomes clear in advance. The individual ergodic theorem and the strong law of large numbers then round out the study of stationary processes.

The second part of the book is mainly about Markov processes; if desired this can be read before most of part one by going directly from Chapter 2 to Chapter 6. I think that the application of semigroup theory is the key to understanding in this area, and so Chapter 7 in which this tool is developed is basic to the discussion. Properties of path functions, strong Markov processes and a little martingale theory are introduced in the later chapters of this part.

I believe that, in the last analysis, probability cannot be properly understood just as pure mathematics, separated from the body of experience and examples which have brought it to life. The students attending my Aarhus lectures already had a considerable acquaintance with applied probability and statistics, and I regard some such experience as one of the essential prerequisites for reading this book with profit. The other prerequisites are a general knowledge of "real analysis" as well as some familiarity with the measure-theoretic formulation of probability theory itself; details are given below. I hope that after finishing this book readers will be prepared either to go on to the frontiers of mathematical research through more specialized literature, or to turn toward applied problems with an ability to relate them to the general theory and to use its tools and ideas as far as may be possible.

If it is true that the mathematics discussed in this book is applicable, the question naturally must arise "Applicable for what?" In the preface to the mimeographed version of my Aarhus lectures,*) given while the Indochina War was still raging, I said:

> "It is impossible for me these days to write or lecture about mathematics without ambivalence. It is obvious that in many nations, and most of all in my own, science and mathematics are all too often serving as tools for militarism and oppression. Probability theory has played a considerable role in some of these perversions, and those who, like myself, work in "pure mathematics" rather than directly with applications must also accept a share of the responsibility. I believe that today it is a vital duty for the scientific community to struggle against such misuse of science, and to resist the demands — made in the name of "defense" or "security" — to develop ever more efficient means for killing and exploiting other human beings."

Such concerns, of course, are not new. The American mathematician who has contributed most to the theories developed in this book is undoubtedly Norbert Wiener. In 1947, in the Introduction to his influential book <u>Cybernetics</u>,**) Wiener wrote:

> "Those of us who have contributed to the new science of cybernetics thus stand in a moral position which is, to say the least, not very comfortable. We have contributed to the initiation of a new science which, as I have said, embraces technical developments with great possibilities for good and for evil. We can only hand it over into the world that exists about us, and this is the world of Belsen and Hiroshima. We do not even have the choice of suppressing these new technical developments. They belong to the age, and the most any of us can do by suppression is to put the development of the subject into the hands of the most irresponsible and most venal of our engineers. The best we can do is to see that a large public understands the trend and the bearing of the present work, and to confine our personal efforts to those fields, such as physiology and psychology, most remote from war and exploitation."

---

*) (Aarhus University Mathematics Lecture Note series, no. 38.)

**) [W], page 28.

That was an important statement, but we must now go further. I believe that scientists have an obligation to try to estimate which of the <u>possible</u> results of new technical developments are <u>likely</u> to occur in reality. This cannot be done in a social and political vacuum. In a peaceful, liberated, nonexploitative society there would be little to fear; beneficial applications would be pushed while harmful ones would wither. But in today's United States it is mainly the government, especially the Pentagon, and the giant corporations which have the resources and the desire to exploit advanced technology for their own purposes. I do not think the prospects here for the benign application of science are encouraging. Elsewhere in the world the outlook is rarely much better, and sometimes worse.

What then can be done? To personally abstain from immediately harmful work is a first step, but no more. Wiener's emphasis on public education is surely important; the vital decisions must not be left to the experts and rulers, but should be made in a broad political forum. This is beginning to happen in the nuclear energy controversy, for example, despite powerful efforts to exclude the public from meaningful participation. Individual scientists and engineers, and several <u>organizations</u> of scientists, have played important roles in this process.

Perhaps the key word which must be added to Wiener's statement is "organize." The great day of the dedicated solitary researcher is over, if indeed it ever existed. Now our scientific work is elaborately planned and supported, but the old individualistic idealogy of "disinterested research" and "knowledge for its own sake" persists. These concepts can serve as intellectual blinders which prevent us from understanding the social role which we in fact do play as mathematicians, scientists and engineers, and which keep us from working effectively for change. In their stead, concern for the human consequences of scientific and technological achievement must become part of our working lives, of our teaching and learning, of our professional meetings and writing. Only through organized collective action can this be achieved.

The goal of controlling and humanizing science will not be fully attained, I believe, until radical changes have been made in the structure of society. I also believe that to wait for that day before beginning to act invites disaster. Fortunately there appear to be a growing number of people, in the U. S. and elsewhere, who are deeply concerned about the social consequences of their scientific work, who are ready to give this concern a major role in their professional lives, and who are getting together in old and new ways to develop their ideas and to put them into practice. Since this must be the starting point, perhaps there is some basis for optimism.

The author of a book such as this one is obviously indebted to almost everyone who has contributed to the field, and I am drawing not only on the research but also on the expository writing of many others. In particular, lecture notes of A. D. Wentzel and (especially) of K. Ito have been very helpful in developing my feeling for stochastic processes, and the writings of Noam Chomsky, D. F. Fleming and

I. F. Stone have aided me to better understand the world in which we live. Other resources are listed in the bibliography.

I wish to express my appreciation for the hospitality of the Mathematics Institute of Aarhus University where my lectures were given four years ago; my visit there was both pleasant and profitable. And finally, I am grateful to the Dartmouth College mathematics department for its generosity with leave and assistance during the preparation of the final version of the book.

John Lamperti
Hanover, N. H.

February, 1977

PREREQUISITES

There are three general mathematical prerequisites for reading this book with profit. They are an adequate knowledge of mathematical analysis, knowledge of basic probability mathematics (including its measure-theoretic foundations), and familiarity with examples and applications from elementary probability, preferably including finite Markov chains.

Taking up the last point first, there is no use in trying to prescribe exactly what must be known; the idea is that readers should have some feeling for the importance of the mathematics we will be discussing, as well as some basic intuition about how probability works. There are innumerable valid ways to gain such experience, but in my opinion anyone who does not yet have it should postpone reading this book. I can think of no better source to which to turn than William Feller's beautiful text [F_1]. The chapters on Markov chains and on examples of continuous-time stochastic processes (chapters 15 and 17 respectively in the 3rd edition) will be especially helpful in motivating most of part 2 here.

It is assumed that the reader has already studied the modern formulation of probability theory: probability spaces, random variables as measurable functions, the concept of independence and some properties of sums of independent random variables such as the laws of large numbers (weak and strong) and the central limit theorem. One source for all this is chapters 1 through 3 of my previous book [L]. Of course there are many other places to find this material, and some of them are listed in the bibliographies here and in [L].

One specific comment: the general (non-discrete) theory of conditional probabilities and expectations is not treated in [L] but will be needed here. For this reason the essentials have been set out in Appendix 2, and any reader who is not familiar with these topics should begin there.

The prerequisites in analysis begin, of course, with an adequate knowledge of measure theory. This must include familiarity with abstract measure spaces and integrals on them, plus such basic results as the extension theorem for constructing σ-additive measures on Borel fields, the dominated and monotone convergence theorems, the Fubini theorem for product measures, and the Radon-Nikodym theorem. In addition to measure theory, considerable use is made of Hilbert and Banach spaces, both in the abstract and such more specific ones as L_2 and spaces of continuous or bounded functions. The concept of "Hilbert space" is not defined in this book; it is assumed that the reader has studied Hilbert spaces already and is familiar with such things as orthogonal bases and series (generalized Fourier series), subspaces, bounded linear operators and functionals, and projections. The spectral theorem is not assumed, although a form of it is basic for Chapters 3 and 4; we will derive what we need there in (I hope) a relatively painless way. As for Banach spaces, the concepts of linear functional and of bounded and unbounded linear operators will be used (in part 2) without special explanations. The results of F. Riesz relating linear functionals on spaces of continuous functions to integrals should also be familiar.

The material above is usually included in a course in "real analysis" such as the one taught at Dartmouth College for first-year mathematics graduate students. All that I have mentioned, and much more, can be found in Rudin's Real and Complex Analysis [R]. I might add, however, that it is certainly not necessary to know everything in that book in order to read this one! Also, I personally feel that maintaining strict logical priorities can be a block to learning mathematics; it is not a sin to understand the statement of a theorem and to use it before learning the proof! If this point of view is accepted, the above list of prerequisities may seem much less formidable.

Some of the more particular bits of information needed are listed below, arranged according to the chapter in which they occur:

Chapter 2. In section 3 some facts about the multivariate normal distribution are used without proof or much explanation; adequate background can be found in [F_2], Chapter 3, section 6.

Chapter 3. When reading section 1 it will help to have seen the solution of linear, homogeneous finite-difference equations. Section 3 requires the use of "Helly's theorems" about the weak convergence of measures on R^1; one reference among many is [L], section 12. Weak convergence turns up elsewhere in the book too.

Chapter 4. Some standard Hilbert-space theory and harmonic analysis are needed throughout the chapter, and in section 2 we need to know that linear functionals on continuous functions are represented by at most one signed measure. For this, see [R], Chapter 6 (the Riesz representation theorem). Then in section 3 we require a theorem of F. and M. Riesz whose proof is sketched later in section 7. That proof (which can be omitted without loss of continuity) involves some complex analysis and a few facts about harmonic functions which go beyond what is needed elsewhere; they are all in [R].

Chapter 7. Basic ideas about Banach spaces are used throughout the chapter, but semigroup theory is developed from scratch. A little advance familiarity with Laplace transforms may help; it's not strictly required. The weak topology is mentioned several times but never really used except at the end of section 6 (which can be omitted). Some knowledge of elementary differential equations will help with certain examples.

Chapters 8 and 9 involve lots of measure theory, but nothing exotic. The last section of Chapter 9 probably can't be appreciated without some prior knowledge of potential theory and harmonic functions (see [R] and/or [K]) but it plays no role in what follows.

And that's about all.

REMARKS ON NOTATION

As a rule, bold-face letters have been used for operators throughout the book; at times they have also been used for the names of particular Hilbert or Banach spaces. The Hardy symbols o and 0 appear occasionally; the statement that a function $f(h) = o(h)$ as $h \to a$ means that $\lim_{h\to a} f(h)/h = 0$, while the similar statement with $O(h)$ means that $f(h)/h$ is bounded. (The "as $h \to a$" may be omitted if it is clear from the context.) A few other symbols used repeatedly, and their meanings, are listed below:

$(\Omega, \mathcal{F}, P)$ is the standard notation for a probability space, consisting of a set Ω, a σ-field of its subsets $\mathcal{F}$, and a probability measure P. The letter E is always used to denote mathematical expectation of a random variable.

$\mathbb{B}(W)$, where W is either a collection of sets or of random variables, means the smallest σ-field containing all the sets in W, or with respect to which all the random variables are measurable, whichever is appropriate.

$\phi_A(x)$ is the indicator function of the set A; that is, $\phi_A(x) = 1$ if $x \in A$ and $= 0$ otherwise.

$\mathbb{C}$ denotes the complex numbers.

R^n is real Euclidean n-space; R^{n*} is real n-space compactified by adding a single "point at ∞."

δ_{ij} is the Kronecker δ, which is 1 if $i = j$ and 0 if $i \neq j$.

$\mathbb{1}$ means the constant function whose value is always 1.

'iff' is sometimes used in place of 'if and only if.'

TABLE OF CONTENTS

CHAPTER 1

GENERAL INTRODUCTION

1. Basic definitions.

A stochastic (or random) process is formally defined to be a collection of random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$ and indexed by the elements of a parameter set T. The set T will in this book generally be one of these:

$$R^1,\ R^+ = [0,\infty),\ \mathbb{Z} = \{\ldots,-1,0,1,2,\ldots\},\ \text{or}\ \ \mathbb{Z}^+ = \{0,1,2,\ldots\};$$

in all these cases the parameter $t \in T$ may usually be thought of as time. If $T = \mathbb{Z}$ or $\mathbb{Z}^+$, one sometimes speaks of a random sequence. If $T = R^n$ with $n > 1$ the process is often called a random field. The random variables of the process need not always be real-valued but must have the same range-space S; this may be R^n (a vector-valued process) or some other measurable space. In the first part of this book the (common) range-space of the random variables making up the process will almost always be the real or complex numbers;

this is not generally true in the latter part, although the reader can privately make that restriction if desired. In any case the range S of the variables is called the state space of the process.

In describing a stochastic process as we have just done there is a certain psychological bias: one tends to regard the process primarily as a function on T whose values for each $t \in T$ are random variables. Of course we are really dealing with one function of two variables, say $X = X(t,\omega)$, where $t \in T$, $\omega \in \Omega$, and where for each fixed t the function $X(t,\cdot)$ is measurable with respect to $\mathcal{F}$. If instead of t we fix an $\omega \in \Omega$, we obtain a function $X(\cdot,\omega): T \to R^1$ (or into whatever the state space S may be) which is called a trajectory or a path-function or sample-function of the process. It is also legitimate, and sometimes most appropriate, to think of the process X as a single random variable whose range is a space of functions on T; the term random function perhaps suggests this point of view.

Let $t_1,t_2,\ldots,t_n \in T$, and let C be any measurable set in S^n. Then the definition

$$P_{t_1,\ldots,t_n}(C) = P(\{\omega \in \Omega: (X(t_1,\omega),\ldots,X(t_n,\omega)) \in C\}) \quad (1)$$

makes sense since $X(t_i,\cdot)$ is $\mathcal{F}$-measurable, and the set function $P_{t_1,\ldots,t_n}(\cdot)$ is a probability measure on the measurable sets of S^n. The measures so obtained, for all finite subsets of T, are called the finite-dimensional distributions of the process. This family of distributions is often (but not always!) the most important aspect of the process, and one frequently needs to study a random process by starting with