RANDOM PROCESSES

MURRAY ROSENBLATT

UNIVERSITY TEXTS IN THE MATHEMATICAL SCIENCES

Random Processes

MURRAY ROSENBLATT
Professor of Probability and Statistics
Division of Applied Mathematics
Brown University

New York
OXFORD UNIVERSITY PRESS
1962

Library of Congress Catalogue Card Number: 61-13568
Printed in the United States of America

ACKNOWLEDGEMENT

I am indebted to D. Rosenblatt who encouraged me to write an introductory book on random processes. He also motivated much of my interest in functions of Markov chains. My thanks are due to my colleagues W. Freiberger and G. Newell who read sections of the manuscript and made valuable suggestions. I would especially like to acknowledge the help of J. Hachigian and T. C. Sun, who looked at the manuscript in some detail and made helpful comments on it. Thanks are due to Ezoura Fonseca for patient and helpful typing. This book was written with the support of the Office of Naval Research.

CONTENTS

RANDOM PROCESSES

NOTATION

$A \cup B$	:	the set of points belonging to either of the sets A and B, usually called the union of A and B.
$\bigcup_i A_i$	:	the set of points belonging to any of the sets A_i.
AB or $A \cap B$	:	the set of points belonging to both of the sets A and B, usually called the product or intersection of the sets A and B.
$\bigcap_i A_i$	:	the set of points belonging to all the sets A_i.
$A - B$	:	the set of points in A but not in B, usually called the difference of the sets A and B.
$A \ominus B$	:	the set of points in A or B but not both, usually called the symmetric difference of the sets A and B.
$x \in A$	:	x an element of the set A.
o	:	$f(x) = o(g(x))$ as $x \to r$ if $\lim_{x \to r} f(x)/g(x) = 0$
O	:	$f(x) = O(g(x))$ as $x \to r$ if $\lvert f(x)/g(x) \rvert \leq K < \infty$ as $x \to r$.
$\approx$	:	$f \approx g$ f is approximately the same as g.
$\cong$	:	$f(x) \cong g(x)$ as $x \to r$ if $\lim_{x \to r} f(x)/g(x) = 1$.
$x \to y+$	:	x approaches y from the right.
$x \bmod r$ with $r > 0$	:	$x \bmod r = x - mr$ where mr is the largest multiple of r less than or equal to x.
$\delta_{\lambda,\mu}$ (Kronecker delta)	:	$\delta_{\lambda,\mu}$ is equal to one if $\lambda = \mu$ and zero otherwise.
Re a	:	real part of the complex number a.
$\{\alpha \mid \ldots\}$	:	the set of α satisfying the condition written in the place indicated by the three dots.

All formulas are numbered starting with (1) at the beginning of each section of each chapter. If a formula is referred to in the same section in which it appears, it will be referred to by number alone. If the formula appears in the same chapter but not in the same section, it will be referred to by number and letter of the section in which it appears. A formula appearing in a different chapter will be referred to by chapter, letter of section, and number. Suppose we are reading in section b of Chapter III. A reference to formula (13) indicates that the formula is listed in the same chapter and section. Formula (a.13) is in section a of the same chapter. Formula (II.a.13) is in section a of Chapter II.

I

INTRODUCTION

This text has as its object an introduction to elements of the theory of random processes. Strictly speaking, only a good background in the topics usually associated with a course in Advanced Calculus (see, for example, the text of Apostol [1]) and the elements of matrix algebra is required although additional background is always helpful. Nonetheless a strong effort has been made to keep the required background on the level specified above. This means that a course based on this book would be appropriate for a beginning graduate student or an advanced undergraduate.

Previous knowledge of probability theory is not required since the discussion starts with the basic notions of probability theory. Chapters II and III are concerned with discrete probability spaces and elements of the theory of Markov chains respectively. These two chapters thus deal with probability theory for finite or countable models. The object is to present some of the basic ideas and problems of the theory in a discrete context where difficulties of heavy technique and detailed measure theoretic discussions do not obscure the ideas and problems. Further, the hope is that the discussion in the discrete context will motivate the treatment in the case of continuous state spaces on intuitive grounds. Of course, measure theory arises quite naturally in probability theory, especially so in areas like that of ergodic theory. However, it is rather extreme and in terms of motivation rather meaningless to claim that probability theory is just measure theory. The basic measure theoretic tools required for discussion in continuous state spaces are introduced in Chapter IV without proof and motivated on intuitive grounds and by comparison with the discrete case. For otherwise, we would get lost in the detailed derivations of measure theory. In fact, throughout the book the presentation is made with the main object understanding of the material on intuitive grounds. If rigorous proofs are proper and meaningful with this view in mind they are presented. In a number of places where such rigorous discussions are too lengthy and do not give much immediate understanding, they may be deleted with heuristic discussions given in their place. However, this will be indicated in the derivations. Attention has been paid to the question of motivating the material in terms of the situations in which the probabilistic problems dealt with typically arise.

The principal topics dealt with in the following chapters are strongly and weakly stationary processes and Markov processes. The basic result in the chapter on strongly stationary processes is the ergodic theorem. The related concepts of ergodicity and mixing are also considered. Fourier analytic methods are the appropriate tools for weakly stationary processes. Random harmonic analysis of these processes is considered at some length in Chapter VII. Associated statistical questions relating to spectral estimation for Gaussian stationary processes are also discussed. Chapter VI deals with Markov processes. The two extremes of jump processes and diffusion processes are dealt with. The discussion of diffusion processes is heuristic since it was felt that the detailed sets of estimates involved in a completely rigorous development were rather tedious and would not reward the reader with a degree of understanding consonant with the time required for such a development.

The topics in the theory of random processes dealt with in the book are certainly not fully representative of the field as it exists today. However, it was felt that they are representative of certain broad areas in terms of content and development. Further, they appeared to be most appropriate for an introduction. For extended discussion of the various areas in the field, the reader is referred to Doob's treatise [12] and the excellent monographs on specific types of processes and their applications.

As remarked before, the object of the book is to introduce the reader as soon as possible to elements of the theory of random processes. This means that many of the beautiful and detailed results of what might be called classical probability theory, that is, the study of independent random variables, are dealt with only insofar as they lead to and motivate study of dependent phenomena. It is hoped that the choice of models of random phenomena studied will be especially attractive to a student who is interested in using them in applied work. One hopes that the book will therefore be appropriate as a text for courses in mathematics, applied mathematics, and mathematical statistics. Various compromises have been made in writing the book with this in mind. They are not likely to please everyone. The author can only offer his apologies to those who are disconcerted by some of these compromises.

Problems are provided for the student. Many of the problems may be nontrivial. They have been chosen so as to lead the student to a greater understanding of the subject and enable him to realize the potential of the ideas developed in the text. There are references to the work of some of the people that developed the theory discussed. The references are by no means complete. However, I hope they do give some sense of historical development of the ideas and techniques as they exist today. Too often, one gets the impression that a body of theory has arisen instantaneously since the usual reference is given to the latest or most current version of that theory. References are also given to more extended developments of theory and its application.

Some of the topics chosen are reflections of the author's interest. This is perhaps especially true of some of the discussion on functions of Markov chains and the uniform mixing condition in Chapters III and VIII. The section on functions of Markov chains does give much more insight into the nature of the Markov assumption. The uniform mixing condition is a natural condition to introduce if one is to have asymptotic normality of averages of dependent processes.

II

BASIC NOTIONS FOR FINITE AND DENUMERABLE STATE MODELS

a. Events and Probabilities of Events

Let us first discuss the intuitive background of a context in which the probability notion arises before trying to formally set up a probability model. Consider an experiment to be performed. Some event A may or may not occur as a result of the experiment and we are interested in a number $P(A)$ associated with the event A that is to be called the probability of A occurring in the experiment. Let us assume that this experiment can be performed again and again under the same conditions, each repetition independent of the others. Let N be the total number of experiments performed and N_A be the number of times event A occurred in these N performances. If N is large, we would expect the probability $P(A)$ to be close to N_A/N

$$P(A) \approx N_A/N. \tag{1}$$

In fact, if the experiment could be performed again and again under these conditions without end, $P(A)$ would be thought of ideally as the limit of N_A/N, as N increases without bound. Of course, all this is an intuitive discussion but it sets the framework for some of the basic properties one expects the probability of an event in an experimental context to have. Thus $P(A)$, the probability of the event A, ought to be a real number greater than or equal to zero and less than or equal to 1

$$0 \leq P(A) \leq 1. \tag{2}$$

Now consider an experiment in which two events A_1, A_2 might occur. Suppose we wish to consider the event "either A_1 or A_2 occurs," which we shall denote notationally by $A_1 \cup A_2$. Suppose the two events are disjoint in the following sense: the event A_1 can occur and the event A_2 can occur but both cannot occur simultaneously. Now consider repeating the same experiment independently a large number of times, say N.

Then intuitively

$$P(A_1) \approx N_{A_1}/N, \qquad P(A_2) \approx N_{A_2}/N,$$
$$P(A_1 \cup A_2) \approx N_{A_1 \cup A_2}/N. \tag{3}$$

But $N_{A_1 \cup A_2}$, the number of times "A_1 or A_2 occurs" in the experiment is equal to $N_{A_1} + N_{A_2}$. Thus if A_1, A_2 are disjoint we ought to have

$$P(A_1 \cup A_2) = P(A_1) + P(A_2). \tag{4}$$

By extension, if a finite number of events $A_1, \ldots, A_n$ can occur in an experiment, let $A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i$ denote the event "either A_1 or A_2 or . . . or A_n occurs in the experiment." If the events are disjoint, that is, no two can occur simultaneously, we anticipate as before that

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i). \tag{5}$$

Of course, if the events are not disjoint such an additivity relation will not hold. The notation $\bigcup A_i$ need not be restricted to a finite collection of events $\{A_i\}$. It will also be used for infinite collections of events. Relation (5) would be expected to hold for a denumerable or countable collection $A_1, A_2, \ldots$ of disjoint events.

There is an interesting but trivial event Ω, the event "something occurs." It is clear that $N_\Omega = N$ and hence

$$P(\Omega) = 1. \tag{6}$$

With each event A there is associated an event $\bar{A}$, "A does not occur." We shall refer to this event as the complement of A. Since $N_{\bar{A}} = N - N_A$ it is natural to set

$$P(\bar{A}) = 1 - P(A). \tag{7}$$

Notice that the complement of Ω, $\phi = \bar{\Omega}$ ("nothing occurs") has probability zero

$$P(\phi) = 1 - P(\Omega) = 0. \tag{8}$$

Let us now consider what is implicit in our discussion above. A family of events is associated with the experiment. The events represent classes of outcomes of the experiment. Call the family of events A associated with the experiment $\mathfrak{F}$. The family of events $\mathfrak{F}$ has the following properties:

1 1. *If the events* $A_1, A_2 \epsilon \mathfrak{F}$ *then the event* $A_1 \cup A_2$, "either A_1 or A_2 occurs," *is an element of* $\mathfrak{F}$.

2. *The event* Ω, "something occurs" *is an element of* $\mathfrak{F}$.
3. *Given any event* $A \epsilon \mathfrak{F}$, *the complementary event* $\bar{A}$, "A does not occur," *is an element of* $\mathfrak{F}$.

Further, a function of the events $A \epsilon \mathfrak{F}$, $P(A)$, is given with the following properties:

2 1. $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ *if* $A_1, A_2 \epsilon \mathfrak{F}$ are disjoint.

Notice that the relation

$$P(\bar{A}) = 1 - P(A) \tag{9}$$

follows from 2.2 and 2.3.

In the case of an experiment with a finite number of possible elementary outcomes we can distinguish between compound and simple events associated with the experiment. A simple event is just the specification of a particular elementary outcome. A compound event is the specification that one of several elementary outcomes has been realized in the experiment. Of course, the simple events are disjoint and can be thought of as sets, each consists of one point, the particular elementary outcome each corresponds to. The compound events are then sets each consisting of several points, the distinct elementary outcomes they encompass. In the probability literature the simple events are at times referred to as the "sample points" of the probability model at hand. The probabilities of the simple events, let us say $E_1, E_2, \ldots, E_n$, are assumed to be specified. Clearly

$$0 \leq P(E_i) \leq 1 \tag{10}$$

and since the simple events are disjoint and exhaustive (in that they account for all possible elementary outcomes of the experiment)

$$\sum_{i=1}^{n} P(E_i) = 1. \tag{11}$$

The probability of any event A by 2.3 is

$$P(A) = \sum_{E_i \subset A} P(E_i). \tag{12}$$

The events A of $\mathfrak{F}$ are the events obtained by considering all possible collections of elementary occurrences. Thus the number of distinct events A of $\mathfrak{F}$ are 2^n altogether. A collection of events (or sets) satisfying conditions 1.1–1.3 is commonly called a *field*. In the case of experiments with an infinite number of possible elementary outcomes one usually wishes to strengthen assumption *1* in the following way:

1 1′. *Given any denumerable* (finite or infinite) *collection of events* $A_1, A_2, \ldots$ of $\mathfrak{F}$ $A_1 \cup A_2 \cup \cdots = \bigcup A_i$ "either A_1 or A_2 or . . . occurs" *is an element of* $\mathfrak{F}$. Such a collection of events or sets with property *1*.1 replaced by *1*.1′ is called a *sigma-field*. In dealing with P as a function of events A of a *σ-field* $\mathfrak{F}$, assumption *2*.3 is strengthened and replaced by

$$2.3' \; P(\bigcup A_i) = \sum_i P(A_i) \text{ if } A_1, A_2, \ldots, \in \mathfrak{F} \tag{13}$$

is a denumerable collection of disjoint events. This property is commonly referred to as countable additivity of the P function.

By introducing "sample points" we are able to speak alternatively of events or sets. In fact disjointness of events means disjointness of the corresponding events viewed as collections of elementary outcomes of the experiment. Generally, it will be quite convenient to think of events as sets and use all the results on set operations which have complete counterparts in operations on events. In fact the $\cup$ operation on events is simply set addition for the events regarded as sets. Similarly complementation of an event amounts to set complementation for the event regarded as a set.

It is very important to note that our basic notion is that of an experiment with outcomes subject to random fluctuation. A family or field of events representing the possible outcomes of the experiment is considered with a numerical value attached to each event. This numerical value or probability associated with the event represents the relative frequency with which one expects the event to occur in a large number of independent repetitions of the experiment. This mode of thought is very much due to von Mises [57].

Let us now illustrate the basic notions introduced in terms of a simple experiment. The experiment considered is the toss of a die. There are six elementary outcomes of the experiment corresponding to the six faces of the die that may face up after a toss. Let E_i represent the elementary event "i faces up on the die after the toss." Let

$$0 \leq p_i = P(E_i) \leq 1 \tag{14}$$

be the probability of E_i. The probability of the compound event $A = \{\text{an even number faces up}\}$ is easily seen to be

$$P(A) = p_2 + p_4 + p_6. \tag{15}$$

The die is said to be a "fair" die if

$$p_1 = p_2 = \cdots = p_6 = 1/6.$$

Another event or set operation that is of importance can be simply derived from those already considered. Given two events $A_1, A_2 \epsilon \mathfrak{F}$, consider the derived event $A_1 \cap A_2$ "both A_1 and A_2 occur." It is clear that

$$A_1A_2 = A_1 \cap A_2 = \overline{(\bar{A}_1 \cup \bar{A}_2)}. \tag{16}$$

b. Conditional Probability, Independence, and Random Variables

A natural and important question is what is to be meant by the conditional probability of an event A_1 given that another event A_2 has occurred. The events A_1, A_2 are, of course, possible outcomes of a given experiment. Let us again think in terms of a large number N of independent repetitions of the experiment. Let N_{A_2} be the number of times A_2 has occurred and $N_{A_1 \cap A_2}$ the number of times A_1 and A_2 have simultaneously occurred in the N repetitions of the experiment. It is quite natural to think of the conditional probability of A_1 given A_2, $P(A_1|A_2)$, as very close to

$$N_{A_1 \cap A_2}/N_{A_2} = \frac{N_{A_1 \cap A_2}}{N} \Big/ \frac{N_{A_2}}{N} \tag{1}$$

if N is large. This motivates the definition of the conditional probability $P(A_1|A_2)$ by

$$P(A_1|A_2) = P(A_1 \cap A_2)/P(A_2) \tag{2}$$

which is well defined as long as $P(A_2) > 0$. If $P(A_2) = 0$, $P(A_1|A_2)$ can be taken as any number between zero and one. Notice that with this definition of conditional probability, given any $B \epsilon \mathfrak{F}$ (the field of events of the experiment) for which $P(B) > 0$, the conditional probability $P(A|B)$, $A \epsilon \mathfrak{F}$, as a function of $A \epsilon \mathfrak{F}$ is a well-defined probability function satisfying 2.1-2.3. It is very easy to verify that

$$\sum_i P(A|E_i)P(E_i) = P(A) \tag{3}$$

where the E_i's are the simple events of the probability field $\mathfrak{F}$. A similar relation will be used later on to define conditional probabilities in the case of experiments with more complicated spaces of sample points (sample spaces).

The term independence has been used repeatedly in an intuitive and unspecified sense. Let us now consider what we ought to mean by the independence of two events A_1, A_2. Suppose we know that A_2 has occurred. It is then clear that the relevant probability statement about A_1 is a statement in terms of the conditional probability of A_1 given A_2. It would be natural to say that A_1 is independent of A_2 if the conditional probability of A_1 given A_2 is equal to the probability of A_1

$$P(A_1|A_2) = P(A_1), \tag{4}$$

that is, the knowledge that A_2 has occurred does not change our expectation of the frequency with which A_1 should occur. Now

$$P(A_1|A_2) = P(A_1 \cap A_2)/P(A_2) = P(A_1)$$

so that

$$P(A_1 \cap A_2) = P(A_1)P(A_2). \tag{5}$$

Note that the argument phrased in terms of $P(A_2|A_1)$ would lead to the same conclusion, namely relation (5). Suppose a denumerable collection (finite or infinite) of events $A_1, A_2 \ldots$ is considered. We shall say that the collection of events is a collection of independent events if every finite subcollection of events $A_{k_1}, \ldots, A_{k_m}$, $1 \leq k_1 < \cdots < k_m$, satisfies the product relation

$$P(A_{k_1}A_{k_2} \cdots A_{k_m}) = \prod_{i=1}^{m} P(A_{k_i}).$$

It is easy to give an example of a collection of events that are pairwise independent but not jointly independent. Let $\mathfrak{F}$ be a field of sets with four distinct simple events E_1, E_2, E_3, E_4

$$P(E_i) = 1/4, \quad i = 1, \ldots, 4. \tag{6}$$

Let the compound events A_i $i = 1, 2, 3$ be given by

$$A_1 = E_1 \cup E_2$$
$$A_2 = E_1 \cup E_3$$
$$A_3 = E_1 \cup E_4.$$

Then

$$P(A_i) = 1/2 \quad i = 1, 2, 3 \tag{7}$$

while

$$P(A_1A_2) = P(A_1A_3) = P(A_2A_3) = P(E_1) = 1/4. \tag{8}$$

The events A_i are clearly pairwise independent. Nonetheless

$$P(A_1A_2A_3) = P(E_1) = 1/4 \neq P(A_1)P(A_2)P(A_3). \tag{9}$$