PROBABILITY, RANDOM PROCESSES, AND ESTIMATION THEORY FOR ENGINEERS

HENRY STARK ■ JOHN W. WOODS

Henry Stark
John W. Woods

PROBABILITY, RANDOM PROCESSES, AND ESTIMATION THEORY FOR ENGINEERS

PRENTICE-HALL, Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging-in-Publication Data

Stark, Henry. (date)
Probability, random processes, and estimation theory for engineers.

Includes bibliographies and index.
1. Probabilities. 2. Stochastic processes.
3. Estimation theory. I. Woods, John W. (John William).
II. Title.
TA340.S7 1986 519.2 85-16882
ISBN 0-13-711706-X

Editorial/production supervision and
interior design: Marlys Lehmann
Cover design: Ben Santora
Manufacturing buyer: Rhett Conklin

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-711706-X 025

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Prentice-Hall of Southeast Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*
Whitehall Books Limited, *Wellington, New Zealand*

PROBABILITY, RANDOM PROCESSES, AND ESTIMATION THEORY FOR ENGINEERS

To
Alice, Larry, and Rich
and
Harriet, Anne, and Chris

PREFACE

This book grew out of course notes used by us to teach two one-semester courses on probability and random processes at Rensselaer Polytechnic Institute (RPI). The probability course at RPI is required of all students in the Computer and Systems Engineering Program and is a highly recommended elective for many others. Most of the students taking the course are engineering or science majors in their junior year. Seniors and many first-year graduate students take the course for credit as well. The emphasis is on introducing fundamental principles and developing skills to solve problems.

The random processes course is typically taken by first-year graduate students and is designed to give students the needed background to take more advanced courses in communications, signal processing, controls, robotics, large-scale systems, and physics-related phenomena. As can be seen, the material goes considerably beyond elementary input-output relations for linear shift-invariant systems. Our experience has been that the present level of the course also gives the student the mathematical background in stochastic processes to pursue M.S.- or Ph.D.-level research in these fields.

In writing this book, we took an "integrated" view of probability and random processes. We felt that probability and random variables were so closely linked to estimation and decision theory that we made a strong attempt at connecting these subjects. Furthermore, we both shared a strong feeling that describing sophisticated applications of the theory of probability and random processes would be, from a pedagogical viewpoint, beneficial for students. Hence, we have included applications to pattern recognition, linear systems, parameter estimation, controls, and communication theory. We also felt that in the age of the computer, the theory of random discrete-time (space) sequences should be given as much weight as continuous-time (space) waveforms. For this reason we devoted an entire chapter to this theory.

In reviewing our own experience, we found that certain closely related topics were often not covered in first courses because the material was deemed too advanced and was then not covered in later courses because the material was considered too basic. Such is often the case with covariance matrices, maximization of quadratic forms, least-squares estimation, and still others. Consequently, we included two chapters dealing with these topics.

The normal use of this book would be as follows: For a first course in probability at, say, the junior or senior year, a reasonable goal is to cover Chapters 1 through 3 with a little material from Chapter 4. Starred sections are often not covered. Homework assignments range from five to ten problems per week. For a first-year graduate course on probability in which the students have had some prior exposure to the subject, Chapters 1 to 3 could be covered in much less time, leaving time for covering Chapters 4 and 5. The material in Chapters 4 and 5 is essential for the statistical-pattern recognition course at RPI. The Gauss-Markov theorem for estimating unknown parameters from measurements corrupted by noise is discussed at length as is the problem of finding the best one-dimensional subspace for separating two classes of statistical objects with well-defined means and covariances.

Chapters 6 through 10 provide the material for a first course in random processes. Beginning with random sequences (Chapter 6), the remaining chapters cover, respectively: random processes; mean-square calculus; stationary processes and sequences; and advanced estimation theory. In general, there is more material than can be covered in a one-semester course. Unless the class is very mature mathematically, Kalman and Wiener filtering are omitted as is Martingale theory upon a first reading.

When course time is reduced as, for example, it might be in schools using the quarter system, it is important that Chapters 6, 7, and 9 be covered, essentially in that order, before teaching material from Chapters 8 and 10. Ideally, it would be preferable to cover Chapter 8 before Chapter 9. But, as the latter covers the basic input-output relations for linear systems excited by random signals, we suggest that it be taught out of normal sequence if the instructor feels there is a danger of running out of time.

We acknowledge a tremendous debt of gratitude to our teachers and students. Thanks are due to the administration of Rensselaer Polytechnic Institute, which graciously recognized the creation of such a book as a scholarly activity. The excellent typing skills and cheerful demeanor of Priscilla

Magilligan were crucial to this effort. One of the authors, Henry Stark, is also very grateful to Peggy and Mark Curchack for providing a warm and comfortable home environment while he worked on the book while on leave at the University of Pennsylvania.

Finally, a project like this can be completed only with the cooperation of the authors' spouses. To Alice and Harriet, we extend our gratitude.

HENRY STARK

JOHN W. WOODS

Troy, New York

CONTENTS

1

INTRODUCTION TO PROBABILITY

1.1 INTRODUCTION: WHY STUDY PROBABILITY?

One of the most frequent questions posed by beginning students of probability is: "Is anything truly random and if so how does one differentiate between the truly random and that which, because of a lack of information, is treated as random but really isn't?" First, regarding the question of truly random phenomena: "Do such things exist?" A theologian might state the case as follows: "We cannot claim to know the Creator's mind, and we cannot predict His actions because He operates on a scale too large to be perceived by man. Hence there are many things we shall never be able to predict no matter how refined our measurements."

At the other extreme from the cosmic scale is what happens at the atomic level. Our friends the physicists speak of such things as the *probability* of an atomic system being in a certain state. The uncertainty principle says that, try as we might, there is a limit to the accuracy with which the position and momentum can be simultaneously ascribed to a particle. Both quantities are fuzzy and indeterminate.

Many, including some of our most famous physicists, believe in an essential randomness of nature. Eugen Merzbacher in his well-known textbook on quantum mechanics [1–1] writes:

> The probability doctrine of quantum mechanics asserts that the indetermination, of which we have just given an example, is a property inherent in nature and not merely a profession of our temporary ignorance from which we expect to be relieved by a future better and more complete theory. The conventional interpretation thus denies the possibility of an ideal theory which would

encompass the present quantum mechanics but would be free of its supposed defects, the most notorious "imperfection" of quantum mechanics being the abandonment of strict classical determinism.

But the issue of determinism versus inherent indeterminism need never even be considered when discussing the validity of the probabilistic approach. The fact remains that there is, quite literally, a nearly uncountable number of situations where we cannot make any categorical deterministic assertion regarding a phenomenon because we cannot measure all the contributing elements. Take, for example, predicting the value of the current $i(t)$ produced by a thermally excited resistor R: Conceivably, we might accurately predict $i(t)$ at some instant t in the future if we could keep track, say, of the 10^{23} or so excited electrons moving in each other's magnetic fields and setting up local field pulses that eventually all contribute to producing $i(t)$. Such a calculation is quite inconceivable, however, and therefore we use a probabilistic model rather than Maxwell's equations to deal with resistor noise. Similar arguments can be made for predicting weather, the outcome of a coin toss, the time to failure of a computer, and many other situations.

Thus to conclude: Regardless of which position one takes, that is, determinism versus indeterminism, we are forced to use probabilistic models in the real world because we do not know, cannot calculate, or cannot measure all the forces contributing to an effect. The forces may be too complicated, too numerous, or too faint.

Probability is a mathematical model to help us study physical systems in an *average sense*. Thus we cannot use probability in any meaningful sense to answer questions such as: "What is the probability that a comet will strike the earth tomorrow?" or "What is the probability that there is life on other planets?"†

R. A. Fisher and R. van Mises, in the first third of the twentieth century, were largely responsible for developing the groundwork of modern probability theory. The modern axiomatic treatment upon which this book is based is largely the result of the work by Andrei N. Kolmogorov [1–2].

1.2 THE DIFFERENT KINDS OF PROBABILITY

There are essentially four kinds of probability. We briefly discuss them here.

A. Probability as Intuition

This kind of probability deals with judgments based on intuition. Thus "She will probably marry him," and "He probably drove too fast," are in this category. A mathematical theory dealing with intuitive probability was developed by B. O. Koopman [1–3]. However, we shall not discuss this subject in this book.

† Nevertheless, certain evangelists deal with this question rather fearlessly, and even a popular astronomer has come up with a figure for this probability. However, whatever probability system these people use, it is not the system that we shall discuss in this book.

B. Probability as the Ratio of Favorable to Total Outcomes (Classical Theory)

In this approach, which is not experimental, the probability of an event is computed *a priori*† by counting the number of ways N_E that E can occur and forming the ratio N_E/N where N is the number of all possible outcomes, that is, the number of all alternatives to E plus N_E. An important notion here is that all outcomes are equally likely. Since equally likely is really a way of saying equally probable, the reasoning is somewhat circular. Suppose we throw a pair of unbiased dice and ask what is the probability of getting a seven? We partition the outcome into 36 equally likely outcomes as shown in Table 1.2–1 where each entry is the sum of the numbers on the two dice.

TABLE 1.2–1 Outcomes of Throwing Two Dice

		1st die					
		1	2	3	4	5	6
2nd die	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

The total number of outcomes is 36 if we keep the dice distinct. The number of ways of getting a seven is $N_7 = 6$. Hence

$$P[\text{getting a seven}] = \tfrac{6}{36} = \tfrac{1}{6}.$$

Example 1.2–1: Throw a fair coin twice (note that since no physical experimentation is involved, there is no problem in postulating an ideal "fair coin"). The possible outcomes are HH, HT, TH, TT. The probability of getting at least one tail T is computed as follows: With E denoting the event of getting at least one tail, the event E is the set of outcomes

$$E = \{HT, TH, TT\}.$$

Thus E occurs whenever the outcome is HT or TH or TT. The number of elements in E is $N_E = 3$; the number of all outcomes, N, is four. Hence

$$P[\text{at least one } T] = \frac{N_E}{N} = \tfrac{3}{4}.$$

The classical theory suffers from at least two significant problems: (1) It cannot deal with outcomes that are not equally likely; and (2) it cannot handle uncountably infinite outcomes without ambiguity (see the example by Athanasios Papoulis [1–4]). Nevertheless, in those problems where it is impractical to actually determine the outcome probabilities by experimentation and where, because of symmetry considerations, one can indeed argue equally likely outcomes the classical theory is useful.

† *A priori* means relating to reasoning from self-evident propositions or presupposed by experience. *A posteriori* means relating to reasoning from observed facts.

Historically, the classical approach was the predecessor of Richard Von Mises' [1–5] relative frequency approach developed in the 1930s.

C. Probability as a Measure of Frequency of Occurrence

The relative-frequency approach to defining the probability of an event E is to perform an experiment n times. The number of times that E appears is denoted by n_E. Then it is tempting to define the probability of E occurring by

$$P[E] = \lim_{n\to\infty} \frac{n_E}{n}. \tag{1.2–1}$$

Quite clearly since $n_E \leq n$, we must have $0 \leq P[E] \leq 1$. One difficulty with this approach is that we can never perform the experiment an infinite number of times so that we can only estimate $P[E]$ from a finite number of trials. Secondly, we *postulate* that n_E/n approaches a limit as n goes to infinity. But consider flipping a fair coin 1000 times. The likelihood of getting exactly 500 heads is very small; in fact, if we flipped the coin 10,000 times, the likelihood of getting exactly 5000 heads is even smaller. As $n \to \infty$, the event of observing exactly $n/2$ heads becomes vanishingly small. Yet our intuition demands that $P[\text{head}] = \frac{1}{2}$ for a fair coin. Suppose we choose a $\delta > 0$; then we shall find experimentally that if the coin is truly fair, the number of times that

$$\left|\frac{n_E}{n} - \frac{1}{2}\right| > \delta \tag{1.2–2}$$

as n becomes large, becomes very small. Thus although it is very unlikely that at any stage of this experiment, especially when n is large, n_E/n is exactly $\frac{1}{2}$, this ratio will nevertheless hover around $\frac{1}{2}$, and the number of times it will make significant excursion away from the vicinity of $\frac{1}{2}$ according to Equation 1.2–2 becomes very small indeed.

Despite these problems with the frequency definition of probability, the relative-frequency concept is essential in applying the probability theory to the physical world.

D. Probability Based on an Axiomatic Theory

This is the approach followed in most modern textbooks on the subject. To develop it we must introduce certain ideas, especially those of a random experiment, a sample description space, and an event. Briefly stated, a random experiment is simply an experiment in which the outcomes are nondeterministic, that is, probabilistic. Hence the word *random* in *random experiment*. The *sample description space* is the set of all outcomes of the experiment. An *event* is a subset of the sample description space that satisfies certain constraints. In general, however, almost any subset of the sample description space is an event.

These notions are refined in the next two sections.