Random Signals and Systems

RICHARD E. MORTENSEN

Random Signals and Systems

RICHARD E. MORTENSEN
University of California at Los Angeles

A WILEY-INTERSCIENCE PUBLICATION
JOHN WILEY & SONS
NEW YORK CHICHESTER BRISBANE TORONTO SINGAPORE

Library of Congress Cataloging in Publication Data:

Mortensen, Richard E.
Random signals and systems.

"A Wiley-Interscience publication."
Bibliography: p.
Includes index.
1. Signal theory (Telecommunication) 2. Stochastic processes. 3. Random variables. I. Title.

TK5102.5M68 1987 621.38'043 86-19007
ISBN 0-471-84364-4

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

This book was written to serve as a textbook for either a senior-level introductory course in stochastic processes or a first-year graduate-level follow-up course. The prerequisites for the senior-level course are courses in mathematics usually required of a senior majoring in electrical engineering, namely, matrix algebra, differential equations, and Laplace and Fourier transforms, plus a one-quarter course in probability. Although no additional specific prerequisites are presumed for the graduate course, it is hoped that graduate students will possess a significant intangible asset sometimes called "mathematical maturity." The courses for which this book serves as a text are prerequisites at UCLA for graduate courses in control systems engineering and communications systems engineering. The book is intended to provide the basic knowledge needed for learning to design analog communication systems and linear control systems operating in an aleatory environment as well as for gaining an understanding of standard digital signal processing techniques.

These courses in stochastic processes in the Electrical Engineering Department at UCLA are the descendants of similar courses that have typically been taught for the past 30 years using as texts such books as Davenport and Root (1958) or Papoulis (1965), which are listed in the reference section. During that time, there have been two historical developments that have had contrary impacts upon the way the subject of stochastic processes now needs to be taught. The first of these is the increasingly high level of mathematical sophistication that one encounters in research papers in the *IEEE Transactions* and many other journals pertinent to the subject. The second is the development and proliferation of the microprocessor and associated software.

A consequence of the first development is that a Ph.D. student in electrical engineering who expects to do cutting-edge research and make a theoretical contribution in a dissertation will have to know measure theory and functional analysis. A consequence of the second development is that recent textbooks for undergraduates on digital filter design make very minimal assumptions concerning the reader's background in calculus, but do presume some familiarity with discrete mathematics. In courses such as the one for which this textbook is designed, we are confronted with the dilemma of taking a student whose undergraduate preparation largely reflects the second trend and attempting to prepare that student for a career in a graduate school curriculum oriented toward the first trend.

Just for the sake of illustration, a recent text that exemplifies the first trend is the excellent book by Wong and Hajek (1985). A contemporary text exemplifying the second trend is the enjoyable book by Williams (1986).

I believe that a student who proposes to do serious work in stochastic processes at the Ph.D. level must take a course in real analysis from the mathematics department, followed by a course in functional analysis using a text such as the one by Balakrishnan (1981). Therefore, I have not endeavored to introduce the student to measure theory in this book. I have attempted to set him or her thinking along appropriate lines, by introducing probability in terms of set theory, by calling the probability set function a "probability measure," and by dropping hints here and there. For the same reason, there is no mention, except here, of the Ito stochastic calculus and associated topics in this book. On the other hand, I have ventured to discuss Hilbert space openly and without embarrassment. Although this choice may appear largely idiosyncratic, it was based on my experience in terms of what seems to work and what does not. I do believe it is desirable to introduce the Karhunen–Loeve expansion in a course such as this one, because it is part of the theoretical basis for such successful practical contributions as the Viterbi decoding algorithm. While discussing Karhunen–Loeve, some information on Hilbert space theory is very useful. It also helps in providing an interpretation of the meaning of the innovations process in Kalman filtering theory.

In order to accommodate our computer-oriented undergraduates, Chapter 2 includes a review of the Gaussian distribution in one and two dimensions and an exploration of some of its properties. This incidentally gives me an opportunity to check out the student's ability to do matrix manipulations and to evaluate multiple integrals. Chapter 4 meets the students on their level by discussing finite length random sequences, and in Appendix 2 we provide a computer program that will generate such sequences for students to investigate as they like.

The essential core of the undergraduate course is the material in Chapters 2, 4, 5, and 6. Depending upon the time available and the refractoriness of the students, material from Chapters 3, 9, and 10 can be introduced. The graduate course reviews all of that material and then takes up the discussion of the Hilbert space of second-order random variables from Chapter 1 and the Hilbert space of square integrable functions from Chapter 8, and then proceeds into the presentation of the Karhunen–Loeve expansion. Again depending upon time and opportunity, we can cover the properties of the conditional multidimensional Gaussian density and the introduction to estimation theory from Chapter 3, the state-space theory of dynamic systems from Chapters 7 and 10, and the introduction to Kalman filtering theory in Chapter 11.

The book was deliberately put in the form of a smorgasbord of topics, for maximum flexibility. The style is informal and discursive in order to keep the attention of most students. The theorem–proof format is used only in a few places where it seems particularly desirable to summarize the development and provide a concise statement of results.

Having now tried at some length to explain what I was trying to do, what I think I did, and why I made those particular choices, it is appropriate to express my gratitude to some of my colleagues for facilitating my efforts. I wish to thank Professors A. V. Balakrishnan, Jack Carlyle, and Stephen Jacobsen, all former chairmen of the former Department of System Science, for creating an environment that stimulated the genesis of this book and for their incredible patience with my idiosyncratic behavior. Further thanks go to Professors C. R. Viswanathan and Fred Allen, the former and the current chairman, respectively, of the Department of Electrical Engineering, which largely absorbed the Department of System Science, for creating the nurturing climate that enabled this book to become a reality. Additional gratitude is due my colleagues on the faculties of those two departments, whom I will not mention individually for fear of overlooking someone, for many enjoyable discussions over the years. I am deeply indebted to my students over the past 20 years, who taught me valuable lessons about education.

Finally, my sincerest thanks go to Sophie Spurrier for typing the entire manuscript and enduring the process of making seemingly endless corrections and improvements.

RICHARD E. MORTENSEN

Los Angeles, California
November 1986

Contents

Chapter 6

Chapter 7

Chapter 8

Appendix 2

Chapter 1

Discussion of Probability and Stochastic Processes

Introduction

The purpose of this book is to present some particular topics from the theory of stochastic processes which have found applications in control and communications engineering. The book has been written on the assumption that the reader has already had an introductory course in probability theory. Nevertheless, for a variety of reasons it seems appropriate and useful to begin with a review of that subject.

In this chapter we provide a review of the main ideas from probability theory that will be needed in understanding the material in this book. Beyond that, we will introduce one or two ideas which will probably be new to the reader, such as the Hilbert space of second-order random variables, that also will be handy to have available. Finally, after we define some terms and develop some concepts, we will explain what we hope the student will acquire from studying the material in this book, and provide a brief survey of the task to be undertaken.

In order to do that, we will provide a tentative definition of the term "stochastic process," as well as a brief discussion of certain kinds of stochastic processes which will be encountered again subsequently. The end of the chapter also contains a short statement explaining why the book has been written the way it has.

Probability

It is widely agreed that a good way to study probability theory is to base it on set theory. We will approach the subject from that standpoint. The term

"set" is, in very rigorous treatments, considered to be an undefined concept which includes certain properties that are assumed in the initial axioms upon which the whole subject is based. Intuitively, a set is a collection of objects. In probability theory, these "objects" are *elementary events*. In set theory, the set of all the objects with which one intends to deal is taken as the *universal set*. In probability theory, the universal set is called the *sample space*.

Suppose one does an experiment in which the element of randomness is known to play a role. For example, conduct a survey by selecting some category of people and asking them questions, or make repeated measurements of some physical variable under circumstances where experimental error is known not to be negligible. Such an experiment is sometimes called a *random experiment*. It is not the structure of the experiment that is random; instead, randomness refers to the fact that the outcome cannot be predicted precisely in advance.

The *statistics* of the experiment refers, at the most primitive level, simply to the data itself. On a more refined level, "statistics" also refers to certain properties the data is found to have after subjecting it to some numerical processing. Probability theory is used to analyze such a random experiment. It is used to decide what kind of numerical processing is appropriate for the data and what kind of statements one can make with confidence concerning the statistics. Even more basically, probability theory is used to determine how the experiment should be structured so that one *can* make meaningful statements with confidence.

In performing such an analysis using probability theory, it turns out to be a disadvantage to have a sample space that is too large or too small. Therefore, the choice of sample space is usually tailored to the experiment in question. For example, suppose the experiment is to flip a coin 10 times, and record the outcome of each flip, that is, whether it is heads or tails. A sample space with only two points in it, heads and tails, is too small and is actually not useful. A sample space with infinitely many points in it is certainly large enough. The problem is, it is so large as to be unwieldy, and it may lead one into mathematical distress of a kind that one prefers to avoid if there be a way of avoiding it.

The sample space for the above experiment which turns out to be "just right" is the set of all *binary sequences of length 10*. There are $2^{10} = 1024$ of these, so this sample space contains 1024 points. Each point is an "elementary event," that is, a complete sequence of 10 flips. A single flip is *not* an elementary event.

In doing mathematical probability theory this way, a numerical probability would *first* be assigned to *each elementary event* (each sequence of length 10). The value of the probability assigned to each event must be a

real number between 0 and 1, and the sum of the values over all 1024 points of the sample space must be exactly 1.

At this juncture, we can look at various subsets of the sample space, for example, the subset consisting of all sequences having heads occur on the first flip. The sum of the values of probability over all of the points in this subset is, *by definition*, the probability of getting a head on the first flip. If that number agrees with what you intuitively feel ought to be the case, then you may say that your coin-flipping model is *realistic*. On the other hand, if that is not the value that you think the event of getting a head on the first flip should have, then you must change the probabilities assigned to the elementary events until things come out the way you want them to.

Probability theory will show you how to make calculations from your mathematical model concerning the probabilities of various events. It is up to *you* to take the responsibility for deciding whether or not the model is realistic. If you test it in situations where the correct answer is already known, and the model gives you the correct answer there, then you may feel confident in trusting it in situations where the answer is unknown.

Let us now give some precise mathematical definitions. The fundamental entity that we require in order to use probability theory is a *probability trio* $(\Omega, \mathscr{A}, P)$. The first member of the trio, Ω, is the *sample space*, which may be either finite, countably infinite, or uncountably infinite. The second member of the trio, $\mathscr{A}$, is the *algebra of admissible subsets of* Ω, also called the *algebra of events*. The third member of the trio, P, is the *probability measure* defined on $\mathscr{A}$. That is, P is a set function. Its argument is one of the sets that belongs to $\mathscr{A}$, and its value is a real number between 0 and 1.

If Ω is a finite set, then $\mathscr{A}$ is simply the collection of *all* subsets of Ω, the so-called *power set* 2^{Ω}. If Ω is an infinite set, it is not possible in general to assign a probability to every one of its subsets in a consistent way without encountering mathematical difficulties. Therefore, the family of subsets of Ω to which probabilities are assigned has to be specified. That is what $\mathscr{A}$ is. Its members obey the rules of Boolean algebra with respect to the operations of union, intersection, and complement.

With these agreements in force, the only conditions that the set function P must satisfy in order to be a probability measure are the following:

1. $P(\varnothing) = 0$ where $\varnothing$ = empty set
2. $P(\Omega) = 1$
3. $P(A) \geq 0$ for every A in $\mathscr{A}$
4. If $A_1, A_2, \ldots$ are disjoint members of $\mathscr{A}$, then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

Random Variables

In addition to the function P defined on $\mathscr{A}$, we also consider functions defined on Ω itself. Any such function is called a *random variable*. If the value of the function is a real number, it is called a real random variable; if the value is a complex number it is called a complex-valued random variable; if the value is a vector in R^n, it is called a vector-valued random variable; and so on. It is customary to abbreviate "random variable" by r.v.

If the set Ω is infinite, then in order to avoid mathematical distress we have to ban certain pathological functions. It is very unlikely such a function would arise in most applications, but we will include this restriction for the sake of precision. Let us explain it further.

The class of admissible random variables must agree with our algebra of admissible sets. We will explain what "agree" means for real r.v.'s; the extension to more general r.v.'s is a technicality. If $X(\omega)$ is a real r.v., then we want to discuss the probability that the value of X falls in some interval I of the real line. In order to do that, we have to be dealing with an event. Therefore, define

$$X^{-1}(I) = \{\omega \in \Omega\colon\ X(\omega) \in I\} \tag{1}$$

The symbol $\in$ means "belongs to." It suffices for this condition to consider only the class of semi-infinite intervals of the form $I = (-\infty, a]$, for every real number a. If for each a, the set $X^{-1}(I)$ is a member of $\mathscr{A}$, then X is an admissible r.v.

Under those circumstances, we are assured that the probability $P\{-\infty < X \leq a\}$ of the event that X is less than or equal to a is well defined. We give this probability a special name. Since it is a function of the parameter a, we call it the *distribution function for the r.v. X*. It is denoted by $F_X(a)$. In symbols:

$$F_X(a) = P\{-\infty < X \leq a\} \tag{2}$$

Under appropriate circumstances, the distribution function $F_X(a)$ turns out to be differentiable with respect to the parameter a. This will happen only when the sample space Ω is uncountably infinite. In those cases it is convenient to work with the *probability density function*, defined as the derivative of F_X. It has become a common practice to use the same letter for the argument of this density function as is used to designate the random variable itself. Although this system may be used without confusion by those proficient in the subject, for students trying to master the fundamentals it is misleading and confusing. In this book we will always use a capital

letter for random variables. The parameter in the density function will then be the corresponding lowercase letter.

DEFINITION. Let X be a real random variable having a probability distribution F_X which is differentiable. Denote the derivative by f_X. Then we call f_X the *probability density function for the r.v. X*. In symbols:

$$f_X(x) = \frac{d}{dx} F_X(x) \tag{3}$$

The values of F_X are probabilities, but the values of f_X are not. Probabilities are found by integrating f_X, for example:

$$P\{a \leq X \leq b\} = \int_a^b f_X(x)\, dx \tag{4}$$

It follows directly from (2) that the distribution function F_X for any r.v. X possesses the following four properties:

1. F_X is nondecreasing: $a < b$ implies $F_X(a) \leq F_X(b)$
2. $\lim_{x \to +\infty} F_X(x) = 1$
3. $\lim_{x \to -\infty} F_X(x) = 0$
4. F_X is continuous from the right, that is, at any discontinuity F_X assumes the upper value.

If F_X is piecewise constant, that is, a staircase function consisting of only finite jumps and constant segments, then X is called a discrete r.v. If F_X has no discontinuities whatsoever, then X is called a *continuous* r.v. A general r.v. is sometimes called *mixed*.

Strictly speaking, only continuous r.v.'s with F_X differentiable can possess density functions, although by resorting to the use of δ functions, which is common in engineering practice, even a discrete r.v. can be assigned a density.

Suppose X is a discrete r.v. which assumes only a finite set of possible values $a_1, a_2, \ldots, a_n$, with respective probabilities $p_1, p_2, \ldots, p_n$. Intuitively, we may say that X can be expected to have value a_k a fraction p_k of the time. If we make many different observations of X and average the results, then as the number of observations becomes infinite the sample average will approach the number

$$\mu = \sum_{k=1}^{n} a_k p_k \tag{5}$$

In (5) we have written μ as a sum over the *range of X*, that is, the set of values assumed by X. Conceptually, it is valuable to realize that this same quantity could also be computed by a sum over the sample space Ω, specifically

$$\mu = \sum_{\omega \in \Omega} X(\omega) P\{\omega \in \Omega: X(\omega) = a_k\} \tag{6}$$

The summation in (6) is accomplished by partitioning Ω into disjoint subsets $A_1, A_2, \ldots, A_n$, such that for each k, A_k is the set of ω points for which $X(\omega)$ assumes the same value a_k.

When X is a continuous r.v., the definition (5) generalizes to

$$\mu = \int_{-\infty}^{\infty} x f_X(x)\, dx \tag{7}$$

The expression (6) generalizes into the Lebesgue integral, as defined in measure theory. A discussion of that is beyond the scope of this book.

The quantity given by (5), (6), or (7) is called the *mean* or *expected value* of X. In rigorous treatments, the most satisfactory way of introducing the expected value operator is to base it on a precise version of (6), which we have here written in a symbolic form to try to suggest the underlying concept.

Since we will mainly be concerned with r.v.'s possessing density functions, we will henceforth take (7) as the definition of the mean, without further comment.

Higher moments are defined analogously, whenever the integrals exist:

$$\mu_n = \int_{-\infty}^{\infty} x^n f_X(x)\, dx \tag{8}$$

When considerable work has to be done involving moments, it is useful to make use of the properties of the *characteristic function* $M(u)$, which is just the Fourier transform of the density:

$$M(u) = \int_{-\infty}^{\infty} e^{iux} f(x)\, dx \tag{9}$$

When the moment μ_n exits, it may be found by the formula

$$\mu_n = (-i)^n \frac{d^n}{du^n} M(u)\Big|_{u=0} \tag{10}$$

If the characteristic function is known, then the density may be recovered