Cambridge University Press
0521838606 - An Introduction to Statistical Signal Processing
Robert M. Gray and Lee D. Davisson
Excerpt
More information

1
Introduction

A random or stochastic process is a mathematical model for a phenomenon that evolves in time in an unpredictable manner from the viewpoint of the observer. The phenomenon may be a sequence of real-valued measurements of voltage or temperature, a binary data stream from a computer, a modulated binary data stream from a modem, a sequence of coin tosses, the daily Dow–Jones average, radiometer data or photographs from deep space probes, a sequence of images from a cable television, or any of an infinite number of possible sequences, waveforms, or signals of any imaginable type. It may be unpredictable because of such effects as interference or noise in a communication link or storage medium, or it may be an information-bearing signal, deterministic from the viewpoint of an observer at the transmitter but random to an observer at the receiver.

The theory of random processes quantifies the above notions so that one can construct mathematical models of real phenomena that are both tractable and meaningful in the sense of yielding useful predictions of future behavior. Tractability is required in order for the engineer (or anyone else) to be able to perform analyses and syntheses of random processes, perhaps with the aid of computers. The "meaningful" requirement is that the models must provide a reasonably good approximation of the actual phenomena. An oversimplified model may provide results and conclusions that do not apply to the real phenomenon being modeled. An overcomplicated one may constrain potential applications, render theory too difficult to be useful, and strain available computational resources. Perhaps the most distinguishing characteristic between an average engineer and an outstanding engineer is the ability to derive effective models providing a good balance between complexity and accuracy.

Random processes usually occur in applications in the context of environments or systems which *change* the processes to produce other processes.

The intentional operation on a signal produced by one process, an "input signal," to produce a new signal, an "output signal," is generally referred to as *signal processing*, a topic easily illustrated by examples.

- A time-varying voltage waveform is produced by a human speaking into a microphone or telephone. The signal can be modeled by a random process. This signal might be modulated for transmission, then it might be digitized and coded for transmission on a digital link. Noise in the digital link can cause errors in reconstructed bits, the bits can then be used to reconstruct the original signal within some fidelity. All of these operations on signals can be considered as signal processing, although the name is most commonly used for manmade operations such as modulation, digitization, and coding, rather than the natural possibly unavoidable changes such as the addition of thermal noise or other changes out of our control.
- For digital speech communications at very low bit rates, speech is sometimes converted into a model consisting of a simple linear filter (called an autoregressive filter) and an input process. The idea is that the parameters describing the model can be communicated with fewer bits than can the original signal, but the receiver can synthesize the human voice at the other end using the model so that it sounds very much like the original signal. A system of this type is called a *vocoder*.
- Signals including image data transmitted from remote spacecraft are virtually buried in noise added to them on route and in the front end amplifiers of the receivers used to retrieve the signals. By suitably preparing the signals prior to transmission, by suitable filtering of the received signal plus noise, and by suitable decision or estimation rules, high quality images are transmitted through this very poor channel.
- Signals produced by biomedical measuring devices can display specific behavior when a patient suddenly changes for the worse. Signal processing systems can look for these changes and warn medical personnel when suspicious behavior occurs.
- Images produced by laser cameras inside elderly North Atlantic pipelines can be automatically analyzed to locate possible anomalies indicating corrosion by looking for locally distinct random behavior.

How are these signals characterized? If the signals are random, how does one find stable behavior or structures to describe the processes? How do operations on these signals change them? How can one use observations based on random signals to make intelligent decisions regarding future behavior? All of these questions lead to aspects of the theory and application of random processes.

Courses and texts on random processes usually fall into either of two general and distinct categories. One category is the common engineering approach, which involves fairly elementary probability theory, standard undergraduate Riemann calculus, and a large dose of "cookbook" formulas – often with insufficient attention paid to conditions under which the formulas are valid. The results are often justified by nonrigorous and occasionally mathematically inaccurate handwaving or intuitive plausibility arguments that may not reflect the actual underlying mathematical structure and may not be supportable by a precise proof. While intuitive arguments can be extremely valuable in providing insight into deep theoretical results, they can be a handicap if they do not capture the essence of a rigorous proof.

A development of random processes that is insufficiently mathematical leaves the student ill prepared to generalize the techniques and results when faced with a real-world example not covered in the text. For example, if one is faced with the problem of designing signal processing equipment for predicting or communicating measurements being made for the first time by a space probe, how does one construct a mathematical model for the physical process that will be useful for analysis? If one encounters a process that is neither stationary nor ergodic (terms we shall consider in detail), what techniques still apply? Can the law of large numbers still be used to construct a useful model?

An additional problem with an insufficiently mathematical development is that it does not leave the student adequately prepared to read modern literature such as the many *Transactions of the IEEE* and the journals of the European Association for Signal, Speech, and Image Processing (EURASIP). The more advanced mathematical language of recent work is increasingly used even in simple cases because it is precise and universal and focuses on the structure common to all random processes. Even if an engineer is not directly involved in research, knowledge of the current literature can often provide useful ideas and techniques for tackling specific problems. Engineers unfamiliar with basic concepts such as *sigma-field* and *conditional expectation* will find many potentially valuable references shrouded in mystery.

The other category of courses and texts on random processes is the typical mathematical approach, which requires an advanced mathematical background of real analysis, measure theory, and integration theory. This approach involves precise and careful theorem statements and proofs, and uses far more care to specify precisely the conditions required for a result to hold. Most engineers do not, however, have the required mathematical background, and the extra care required in a completely rigorous development severely limits the number of topics that can be covered in a typical course – in particular, the applications that are so important to engineers tend to be neglected. In addition, too much time is spent with the formal details, obscuring the often simple and elegant ideas behind a proof. Often little, if any, physical motivation for the topics is given.

This book attempts a compromise between the two approaches by giving the basic theory and a profusion of examples in the language and notation of the more advanced mathematical approaches. The intent is to make the crucial concepts clear in the traditional elementary cases, such as coin flipping, and thereby to emphasize the mathematical structure of all random processes in the simplest possible context. The structure is then further developed by numerous increasingly complex examples of random processes that have proved useful in systems analysis. The complicated examples are constructed from the simple examples by signal processing, that is, by using a simple process as an input to a system whose output is the more complicated process. This has the double advantage of describing the action of the system, the actual signal processing, and the interesting random process which is thereby produced. As one might suspect, signal processing also can be used to produce simple processes from complicated ones.

Careful proofs are usually constructed only in elementary cases. For example, the fundamental theorem of expectation is proved only for discrete random variables, where it is proved simply by a change of variables in a sum. The continuous analog is subsequently given without a careful proof, but with the explanation that it is simply the integral analog of the summation formula and hence can be viewed as a limiting form of the discrete result. As another example, only weak laws of large numbers are proved in detail in the mainstream of the text, but the strong law is treated in detail for a special case in a starred section. Starred sections are used to delve into other relatively advanced results, for example the use of mean square convergence ideas to make rigorous the notion of integration and filtering of continuous time processes.

By these means we strive to capture the spirit of important proofs without undue tedium and to make plausible the required assumptions and constraints. This, in turn, should aid the student in determining when certain tools do or do not apply and what additional tools might be necessary when new generalizations are required.

A distinct aspect of the mathematical viewpoint is the "grand experiment" view of random processes as being a probability measure on sequences (for discrete time) or waveforms (for continuous time) rather than being an infinity of smaller experiments representing individual outcomes (called random variables) that are somehow glued together. From this point of view random variables are merely special cases of random processes. In fact, the grand experiment viewpoint was popular in the early days of applications of random processes to systems and was called the "ensemble" viewpoint in the work of Norbert Wiener and his students. By viewing the random process as a whole instead of as a collection of pieces, many basic ideas, such as stationarity and ergodicity, that characterize the dependence on time of probabilistic descriptions and the relation between time averages and probabilistic averages are much easier to define and study. This also permits a more complete discussion of processes that violate such probabilistic regularity requirements yet still have useful relations between time and probabilistic averages.

Even though a student completing this book will not be able to follow the details in the literature of many proofs of results involving random processes, the basic results and their development and implications should be accessible, and the most common examples of random processes and classes of random processes should be familiar. In particular, the student should be well equipped to follow the gist of most arguments in the various *Transactions of the IEEE* dealing with random processes, including the *IEEE Transactions on Signal Processing*, *IEEE Transactions on Image Processing*, *IEEE Transactions on Speech and Audio Processing*, *IEEE Transactions on Communications*, *IEEE Transactions on Control*, and *IEEE Transactions on Information Theory*, and the EURASIP/Elsevier journals such as *Image Communication*, *Speech Communication*, and *Signal Processing*.

It also should be mentioned that the authors are electrical engineers and, as such, have written this text with an electrical engineering flavor. However, the required knowledge of classical electrical engineering is slight, and engineers in other fields should be able to follow the material presented.

This book is intended to provide a one-quarter or one-semester course that develops the basic ideas and language of the theory of random processes and provides a rich collection of examples of commonly encountered processes, properties, and calculations. Although in some cases these examples may seem somewhat artificial, they are chosen to illustrate the way engineers should think about random processes. They are selected for simplicity and conceptual content rather than to present the method of solution to some particular application. *Sections that can be skimmed or omitted for the shorter one-quarter curriculum are marked with a star ($\star$).* Discrete time processes are given more emphasis than in many texts because they are simpler to handle and because they are of increasing practical importance in digital systems. For example, linear filter input/output relations are carefully developed for discrete time; then the continuous time analogs are obtained by replacing sums with integrals. The mathematical details underlying the continuous time results are found in a starred section.

Most examples are developed by beginning with simple processes. These processes are filtered or modulated to obtain more complicated processes. This provides many examples of typical probabilistic computations on simple processes and on the output of operations on simple processes. Extra tools are introduced as needed to develop properties of the examples.

The prerequisites for this book are elementary set theory, elementary probability, and some familiarity with linear systems theory (Fourier analysis, convolution, discrete and continuous time linear filters, and transfer functions). The elementary set theory and probability may be found, for example, in the classic text by Al Drake [18] or in the current MIT basic probability text by Bertsekas and Tsitsiklis [3]. The Fourier and linear systems material can by found in numerous texts, including Gray and Goodman [33]. Some of these basic topics are reviewed in this book in Appendix A. These results are considered prerequisite as the pace and density of material would likely be overwhelming to someone not already familiar with the fundamental ideas of probability such as probability mass and density functions (including the more common named distributions), computing probabilities, derived distributions, random variables, and expectation. It has long been the authors' experience that the students having the most difficulty with this material are those with little or no experience with elementary probability.

Organization of the book

Chapter 2 provides a careful development of the fundamental concept of probability theory – a probability space or experiment. The notions of sample space, event space, and probability measure are introduced and illustrated by examples. Independence and elementary conditional probability are developed in some detail. The ideas of signal processing and of random variables are introduced briefly as functions or operations on the output of an experiment. This in turn allows mention of the idea of expectation at an early stage as a generalization of the description of probabilities by sums or integrals.

Chapter 3 treats the theory of measurements made on experiments: random variables, which are scalar-valued measurements; random vectors, which are a vector or finite collection of measurements; and random processes, which can be viewed as sequences or waveforms of measurements. Random variables, vectors, and processes can all be viewed as forms of signal processing: each operates on "inputs," which are the sample points of a probability space, and produces an "output," which is the resulting sample value of the random variable, vector, or process. These output points together constitute an output sample space, which inherits its own probability measure from the structure of the measurement and the underlying experiment. As a result, many of the basic properties of random variables, vectors, and processes follow from those of probability spaces. Probability distributions are introduced along with probability mass functions, probability density functions, and cumulative distribution functions. The basic derived distribution method is described and demonstrated by example. A wide variety of examples of random variables, vectors, and processes are treated. Expectations are introduced briefly as a means of characterizing distributions and to provide some calculus practice.

Chapter 4 develops in depth the ideas of expectation – averages of random objects with respect to probability distributions. Also called probabilistic averages, statistical averages, and ensemble averages, expectations can be thought of as providing simple but important parameters describing probability distributions. A variety of specific averages are considered, including mean, variance, characteristic functions, correlation, and covariance. Several examples of unconditional and conditional expectations and their properties and applications are provided. Perhaps the most important application is to the statement and proof of laws of large numbers or ergodic theorems, which relate long-term sample-average behavior of random processes to expectations. In this chapter laws of large numbers are proved for simple, but important, classes of random processes. Other important applications of expectation arise in performing and analyzing signal processing applications such as detecting, classifying, and estimating data. Minimum mean squared nonlinear and linear estimation of scalars and vectors is treated in some detail, showing the fundamental connections among conditional expectation, optimal estimation, and second-order moments of random variables and vectors.

Chapter 5 concentrates on the computation and applications of second-order moments – the mean and covariance – of a variety of random processes. The primary example is a form of derived distribution problem: if a given random process with known second-order moments is put into a linear system what are the second-order moments of the resulting output random process? This problem is treated for linear systems represented by convolutions and for linear modulation systems. Transform techniques are shown to provide a simplification in the computations, much like their ordinary role in elementary linear systems theory. Mean square convergence is revisited and several of its applications to the analysis of continuous time random processes are collected under the heading of mean square calculus. Included are a careful definition of integration and filtering of random processes, differentiation of random processes, and sampling and orthogonal expansions of random processes. In all of these examples the behavior of the second-order moments determines the applicability of the results. The chapter closes with a development of several results from the theory of linear least squares estimation. This provides an example of both the computation and the application of second-order moments.

In Chapter 6 a variety of useful models of sometimes complicated random processes are developed. A powerful approach to modeling complicated random processes is to consider linear systems driven by simple random processes. Chapter 5 used this approach to compute second-order moments, this chapter goes beyond moments to develop a complete description of the output processes. To accomplish this, however, one must make additional assumptions on the input process and on the form of the linear filters. The general model of a linear filter driven by a memoryless process is used to develop several popular models of discrete time random processes. Analogous continuous time random process models are then developed by direct description of their behavior. The principal class of random processes considered is the class of independent increment processes, but other processes with similar definitions but quite different properties are also introduced. Among the models considered are autoregressive processes, moving-average processes, ARMA (autoregressive moving-average) processes, random walks, independent increment processes, Markov processes, Poisson and Gaussian processes, and the random telegraph wave process. We also briefly consider an example of a nonlinear system where the output random processes can at least be partially described – the exponential function of a Gaussian or Poisson process which models phase or frequency modulation. We close with examples of a type of "doubly stochastic" process – a compound process formed by adding a random number of other random effects.

Appendix A sketches several prerequisite definitions and concepts from elementary set theory and linear systems theory using examples to be encountered elsewhere in the book. The first subject is crucial at an early stage and should be reviewed before proceeding to Chapter 2. The second subject is not required until Chapter 5, but it serves as a reminder of material with which the student should already be familiar. Elementary probability is not reviewed, as our basic development includes elementary probability presented in a rigorous manner that sets the stage for more advanced probability. The review of prerequisite material in the appendix serves to collect together some notation and many definitions that will be used throughout the book. It is, however, only a brief review and cannot serve as a substitute for a complete course on the material. This chapter can be given as a first reading assignment and either skipped or skimmed briefly in class; lectures can proceed from an introduction, perhaps incorporating some preliminary material, directly to Chapter 2.

Appendix B provides some scattered definitions and results needed in the book that detract from the main development, but may be of interest for background or detail. These fall primarily in the realm of calculus and range from the evaluation of common sums and integrals to a consideration of different definitions of integration. Many of the sums and integrals should be prerequisite material, but it has been the authors' experience that many students have either forgotten or not seen many of the standard tricks. Hence several of the most important techniques for probability and signal processing applications are included. Also in this appendix some background information on limits of double sums and the Lebesgue integral is provided.

Appendix C collects the common univariate probability mass functions and probability density functions along with their second-order moments for reference.

The book concludes with Appendix D suggesting supplementary reading, providing occasional historical notes, and delving deeper into some of the technical issues raised in the book. In that section we assemble references on additional background material as well as on books that pursue the various topics in more depth or on a more advanced level. We feel that these comments and references are supplementary to the development and that less clutter results by putting them in a single appendix rather than strewing them throughout the text. The section is intended as a guide for further study, not as an exhaustive description of the relevant literature, the latter goal being beyond the authors' interests and stamina.

Each chapter is accompanied by a collection of problems, many of which have been contributed by collegues, readers, students, and former students. It is important when doing the problems to justify any "yes/no" answers. If an answer is "yes," prove it is so. If the answer is "no," provide a counterexample.

2
Probability

2.1 Introduction

The theory of random processes is a branch of probability theory and probability theory is a special case of the branch of mathematics known as measure theory. Probability theory and measure theory both concentrate on functions that assign real numbers to certain sets in an abstract space according to certain rules. These set functions can be viewed as measures of the size or weight of the sets. For example, the precise notion of area in two-dimensional Euclidean space and volume in three-dimensional space are both examples of measures on sets. Other measures on sets in three dimensions are mass and weight. Observe that from elementary calculus we can find volume by integrating a constant over the set. From physics we can find mass by integrating a mass density or summing point masses over a set. In both cases the set is a region of three-dimensional space. In a similar manner, probabilities will be computed by integrals of densities of probability or sums of "point masses" of probability.

Both probability theory and measure theory consider only nonnegative real-valued set functions. The value assigned by the function to a set is called the *probability* or the *measure* of the set, respectively. The basic difference between probability theory and measure theory is that the former considers only set functions that are normalized in the sense of assigning the value of 1 to the entire abstract space, corresponding to the intuition that the abstract space contains every possible outcome of an experiment and hence should happen with certainty or probability 1. Subsets of the space have some uncertainty and hence have probability less than 1.

Probability theory begins with the concept of a *probability space*, which is a collection of three items:

1. An *abstract space* Ω, as encountered in Appendix A, called a *sample space*, which