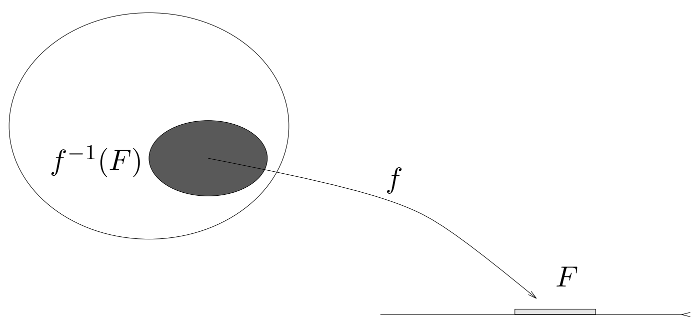
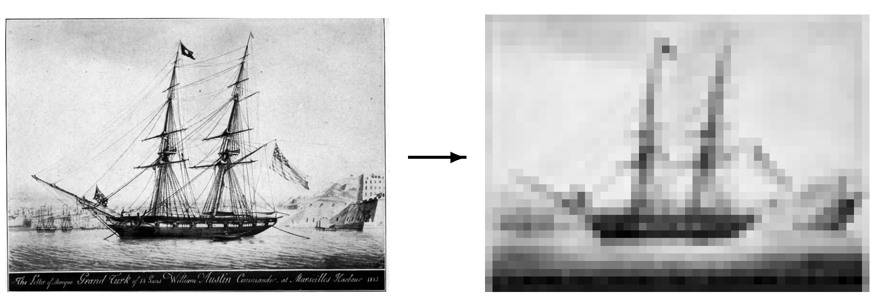
An Introduction to Statistical Signal Processing



$$\Pr(f \in F) = P(\{\omega : \omega \in F\}) = P(f^{-1}(F))$$



January 4, 2011

An Introduction to Statistical Signal Processing

Robert M. Gray
and
Lee D. Davisson

Information Systems Laboratory
Department of Electrical Engineering
Stanford University
and
Department of Electrical Engineering and Computer Science
University of Maryland

to our Families

Contents

Preface

The origins of this book lie in our earlier book *Random Processes: A Mathematical Approach for Engineers* (Prentice Hall, 1986). This book began as a second edition to the earlier book and the basic goal remains unchanged – to introduce the fundamental ideas and mechanics of random processes to engineers in a way that accurately reflects the underlying mathematics, but does not require an extensive mathematical background and does not belabor detailed general proofs when simple cases suffice to get the basic ideas across. In the years since the original book was published, however, it has evolved into something bearing little resemblance to its ancestor. Numerous improvements in the presentation of the material have been suggested by colleagues, students, teaching assistants, and reviewers, and by our own teaching experience. The emphasis of the book shifted increasingly towards examples and a viewpoint that better reflected the title of the courses we taught using the book for many years at Stanford University and at the University of Maryland: *An Introduction to Statistical Signal Processing.* Much of the basic content of this course and of the fundamentals of random processes can be viewed as the analysis of statistical signal processing systems: typically one is given a probabilistic description for one random object, which can be considered as an *input signal.* An operation is applied to the input signal (*signal processing*) to produce a new random object, the *output signal.* Fundamental issues include the nature of the basic probabilistic description, and the derivation of the probabilistic description of the output signal given that of the input signal and the particular operation performed. A perusal of the literature in statistical signal processing, communications, control, image and video processing, speech and audio processing, medical signal processing, geophysical signal processing, and classical statistical areas of time series analysis, classification and regression, and pattern recognition shows a wide variety of probabilistic models for input processes and for operations on those processes, where the operations might be deterministic or random, natural or artificial, linear or nonlinear, digital or analog, or beneficial or harmful. An introductory course focuses on the fundamentals underlying the analysis of such systems: the theories of probability, random processes, systems, and signal processing.

When the original book went out of print, the time seemed ripe to convert the manuscript from the prehistoric troff format to the widely used LaTeX format and to undertake a serious revision of the book in the process. As the revision became more extensive, the title changed to match the course name and content. We reprint the original preface to provide some of the original motivation for the book, and then close this preface with a description of the goals sought during the many subsequent revisions.

Preface to Random Processes: An Introduction for Engineers

> Nothing in nature is random ... A thing appears random only through the incompleteness of our knowledge.
>
> *Spinoza, Ethics I*

> I do not believe that God rolls dice.
>
> *attributed to Einstein*

Laplace argued to the effect that given complete knowledge of the physics of an experiment, the outcome must always be predictable. This metaphysical argument must be tempered with several facts. The relevant parameters may not be measurable with sufficient precision due to mechanical or theoretical limits. For example, the uncertainty principle prevents the simultaneous accurate knowledge of both position and momentum. The deterministic functions may be too complex to compute in finite time. The computer itself may make errors due to power failures, lightning, or the general perfidy of inanimate objects. The experiment could take place in a remote location with the parameters unknown to the observer; for example, in a communication link, the transmitted message is unknown a priori, for if it were not, there would be no need for communication. The results of the experiment could be reported by an unreliable witness – either incompetent or dishonest. For these and other reasons, it is useful to have a theory for the analysis and synthesis of processes that behave in a random or unpredictable manner. The goal is to construct mathematical models that lead to reasonably accurate prediction of the long-term average behavior of random processes. The theory should produce good estimates of the average behavior of real processes and thereby correct theoretical derivations with measurable results.

In this book we attempt a development of the basic theory and applications of random processes that uses the language and viewpoint of rigorous mathematical treatments of the subject but which requires only a typical bachelor's degree level of electrical engineering education including elementary discrete and continuous time linear systems theory, elementary probability, and transform theory and applications. Detailed proofs are presented only when within the scope of this background. These simple proofs, however, often provide the groundwork for "handwaving" justifications of more general and complicated results that are semi-rigorous in that they can be made rigorous by the appropriate delta-epsilontics of real analysis or measure theory. A primary goal of this approach is thus to use intuitive arguments that accurately reflect the underlying mathematics and which will hold up under scrutiny if the student continues to more advanced courses. Another goal is to enable the student who might not continue to more advanced courses to be able to read and generally follow the modern literature on applications of random processes to information and communication theory, estimation and detection, control, signal processing, and stochastic systems theory.

Revisions

Through the years the original book has continually expanded to roughly double its original size to include more topics, examples, and problems. The material has been significantly reorganized in its grouping and presentation. Prerequisites and preliminaries have been moved to the appendices. Major additional material has been added on jointly Gaussian vectors, minimum mean squared error estimation, linear and affine least squared error estimation, detection and classification, filtering, and, most recently, mean square calculus and its applications to the analysis of continuous time processes. The index has been steadily expanded to ease navigation through the book. Numerous errors reported by reader email have been fixed and suggestions for clarifications and improvements incorporated.

This book is a work in progress. Revised versions will be made available through the World Wide Web page http://ee.stanford.edu/˜gray/sp.html. The material is copyrighted by Cambridge University Press, but is freely available as a pdf file to any individuals who wish to use it provided only that the contents of the entire text remain intact and together. Comments, corrections, and suggestions should be sent to rmgray@stanford.edu. Every effort will be made to fix typos and take suggestions into account on at least an annual basis.

Acknowledgements

We repeat our acknowledgements of the original book: to Stanford University and the University of Maryland for the environments in which the book was written, to the John Simon Guggenheim Memorial Foundation for its support of the first author during the writing in 1981–2 of the original book, to the Stanford University Information Systems Laboratory Industrial Affiliates Program which supported the computer facilities used to compose this book, and to the generations of students who suffered through the ever changing versions and provided a stream of comments and corrections. Thanks are also due to Richard Blahut and anonymous referees for their careful reading and commenting on the original book. Thanks are due to the many readers who have provided corrections and helpful suggestions through the Internet since the revisions began being posted. Particular thanks are due to Yariv Ephraim for his continuing thorough and helpful editorial commentary. Thanks also to Sridhar Ramanujam, Raymond E. Rogers, Isabel Milho, Zohreh Azimifar, Dan Sebald, Muzaffer Kal, Greg Coxson, Mihir Pise, Mike Weber, Munkyo Seo, James Jacob Yu, and several anonymous reviewers for Cambridge University Press. Thanks also to Philip Meyler, Lindsay Nightingale, and Joseph Bottrill of Cambridge University Press for their help in the production of the final version of the book. Thanks to the careful readers who informed me of typos and mistakes in the book following its publication, all of which have been reported and fixed in the errata (http://ee.stanford.edu/~gray/sperrata.pdf) and incorporated into the electronic version: Ian Lee, Michael Gutmann, André Isidio de Melo, and especially Ron Aloysius, who has contributed greatly to the fixing of typos. Lastly, the first author would like to acknowledge his debt to his professors who taught him probability theory and random processes, especially Al Drake and Wilbur B. Davenport Jr. at MIT and Tom Pitcher at USC.

Glossary

{ }	a collection of points satisfying some property, e.g. $\{r : r \leq a\}$ is the collection of all real numbers less than or equal to a value a
[ ]	an interval of real points including the end points, e.g. for $a \leq b$ $[a, b] = \{r : a \leq r \leq b\}$. Called a *closed interval*
( )	an interval of real points excluding the end points, e.g. for $a \leq b$ $(a, b) = \{r : a < r < b\}$. Called an *open interval*. Note this is empty if $a = b$
( ], [ )	denote intervals of real points including one endpoint and excluding the other, e.g. for $a \leq b$ $(a, b] = \{r : a < r \leq b\}$, $[a, b) = \{r : a \leq r < b\}$
$\emptyset$	the empty set, the set that contains no points.
$\forall$	for all
Ω	the sample space or universal set, the set that contains all of the points
$\#(F)$	the number of elements in a set F
$\triangleq$	equal by definition
exp	the exponential function, $\exp(x) \triangleq e^x$, used for clarity when x is complicated
$\mathcal{F}$	sigma-field or event space
$\mathcal{B}(\Omega)$	Borel field of Ω, that is, the sigma-field of subsets of the real line generated by the intervals or the Cartesian product of a collection of such sigma-fields
iff	if and only if
l.i.m.	limit in the mean
$o(u)$	function of u that goes to zero as $u \to 0$ faster than u

P	probability measure
P_X	distribution of a random variable or vector X
p_X	probability mass function (pmf) of a random variable X
f_X	probability density function (pdf) of a random variable X
F_X	cumulative distribution function (cdf) of a random variable X
$E(X)$	expectation of a random variable X
$M_X(ju)$	characteristic function of a random variable X
$\oplus$	addition modulo 2
$1_F(x)$	indicator function of a set F: $1_F(x) = 1$ if $x \in F$ and 0 otherwise
Φ	Φ-function (Eq. (2.78))
Q	complementary Phi function (Eq. (2.79))
$\mathcal{Z}_k$	$\triangleq \{0, 1, 2, \ldots, k-1\}$
$\mathcal{Z}_+$	$\triangleq \{0, 1, 2, \ldots\}$, the collection of nonnegative integers
$\mathcal{Z}$	$\triangleq \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the collection of all integers

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