Eugene Wong

Introduction to Random Processes

Consulting Editor: John B. Thomas

Eugene Wong

Introduction to Random Processes

Consulting Editor: John B. Thomas

With 26 Illustrations

Springer-Verlag
New York Heidelberg Berlin

A Dowden &
Culver Book

Eugene Wong
Department of Electrical Engineering
and Computer Sciences
University of California
Berkeley, California 94720
U.S.A.

Library of Congress Cataloging in Publication Data
Wong, Eugene, 1934–
Introduction to random processes.
(Springer texts in electrical engineering)
"A Dowden & Culver book."
Bibliography: p.
Includes index.
1. Stochastic processes. I. Thomas, John Bowman,
1925– . II. Title. III. Series.
QA274.W66 1983 519.2 83-358

Printed and bound by R.R. Donnelley & Sons, Harrisonburg, Va.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-90757-2 Springer-Verlag New York Heidelberg Berlin
ISBN 3-540-90757-2 Springer-Verlag Berlin Heidelberg New York

Preface

The material in this text was developed for a first-year graduate course on stochastic process at Berkeley. While some background on probability theory and some degree of mathematical sophistication are assumed of the student, the book is largely self-contained as to definitions, concepts, and the principal results that are used. Mathematical details are sometimes omitted and these can be found in the references listed at the end of the book.

While the chapters are arranged in their logical order, the core material begins with Chapter 4. Each of the last four chapters (4-7) deals with a major topic in applied stochastic process theory, while the first three chapters deal with mathematical fundamentals. From a pedagogical point of view, some instructors may well prefer to begin with Chapter 4 and to fill in the background material as it is needed. Depending on how much of the final two chapters is included, the topics in this book can be covered in a quarter (30-40 lecture hours) or a semester (45-60 lecture hours). I have done both at Berkeley.

A short collection of exercises can be found at the end of the book. It is hoped that these would serve as prototypes from which additional problems could be developed.

A word on notation. For typing ease the exponential function is sometimes written without raising the exponent, e.g., $e-(\frac{1}{2})x^2$. I know of no instance where this notation is ambiguous.

I am grateful to my former teacher Professor John B. Thomas for encouraging me to put this material into book form, and for reviewing the manuscript. I would also like to thank Ms. Doris Simpson for a skillful job in preparing the camera copy from which this book was produced.

Contents

Chapter 1	Event and Probability	1
1	Introduction	1
2	Computation of Probabilities	2
3	Independent and Conditional Probability	7
Chapter 2	Random Variables	10
1	Definition and Distribution Function	10
2	Expectation	12
3	Finite Collection of Random Variables	15
4	Events Determined by Random Variables	17
5	Functions of Random Variables	21
6	Moments and Characteristic Function	28
7	Independent Random Variables and Conditional Density Functions	31
8	Conditional Expectation	36
9	Gaussian Random Variables	39
Chapter 3	Random Sequences	47
1	Finite Dimensional Distributions	47
2	Convergence Concepts	49
3	Limit Theorems and Sampling	55
Chapter 4	Stochastic Processes	59
1	Introduction	59
2	Continuity Concepts	62
3	Covariance Function	64
4	Gaussian Process and Brownian Motion	67
5	Martingales and Markov Processes	71
6	Stationarity and Time Average	76
7	Orthogonal Representations	81
Chapter 5	Frequency-Domain Analysis	86
1	Introduction	86
2	Fourier Integrals	89
3	Impulse Response and Causal Systems	94
4	Spectral Density	98
5	White Noise Representations	104
6	Sampling Theorem	106
7	Wiener Filtering	108

Chapter 6	Dynamical Systems	115
1	Linear Differential Systems	115
2	Recursive Filtering	118
3	Nonlinear Differential Systems	124
Chapter 7	Likelihood Ratios and Applications	131
1	Likelihood Ratios	131
2	Sequences and Processes	136
3	Hypothesis Testing and Signal Detection	144
4	Additive Gaussian Noise	149
5	Applications in Estimation	152
Suggested Further Readings		157
Exercises		158
Index		174

Chapter 1. Event and Probability

1. Introduction

For most people, intuitive notions concerning probabilities are connected with relative frequencies of occurrence. For example, when we say that in tossing a coin, the probability of its coming up "heads" is 1/2, we usually mean that in a large number of tosses, about 1/2 of the tosses will come up heads. Unfortunately, relative frequency of occurrence has proved to be an unsatisfactory starting point in defining probability. Although there have been attempts to make frequency of occurrence part of the axiomatic structure of probability theory, the currently accepted formulation is one based on measure theory due to Kolmogorov. In this formulation frequency of occurrence is an interpretation for probability rather than a definition. This interpretation is justified under suitable conditions by the law of large numbers.

The starting point of probability theory is usually taken to be an experiment the outcome of which is not fixed a priori. Some familiar examples include tossing a die, observation of a noise voltage at a fixed time, the error in measuring a physical parameter, and the exact touchdown time of an aircraft. Let Ω denote the set of all possible outcomes of an experiment. For example, for the experiment of tossing one die, $\Omega = \{1, 2, 3, 4, 5, 6\}$, while for the touchdown time of an aircraft, Ω might be chosen to be $0 \leq t < \infty$. We note that for a given experiment only one outcome is ever observed. For example, if we toss a die twice, we can consider the first toss as one experiment and the second toss as a separate experiment, or better yet, consider the two tosses together as a single experiment with 36 possible outcomes, each outcome being a pair of numbers (i, j), $i, j = 1, 2, \cdots, 6$. It is better because we can then consider concepts that involve both tosses, e.g., 6 on either throw. In any event, we do not consider the results of the two throws as different outcomes of the same experiment.

Probability is a nonnegative number assigned to certain subsets of Ω. Thus it is a set function, and we shall denote the probability of a set A by $P(A)$. Every probability must satisfy the following three axioms.

Axiom 1 $\qquad P(A) \geq 0, \quad P(\Omega) = 1$

Axiom 2 $\qquad$ If A and B are disjoint sets, i.e., if the intersection $A \cap B$ is empty, then

(additivity) $$P(A \cup B) = P(A) + P(B)$$

Axiom 3 If $A_1, A_2, \cdots,$ is a convergent sequence of sets, then

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

For experiments with only a finite number of outcomes, the third axiom is unnecessary. Axiom 3 is known as sequential continuity. In the next section we explain further the concept of the limit of a sequence of sets.

A subset of Ω for which a probability is defined is called an event. If Ω contains only a finite number of outcomes, or even a countable number of outcomes, then every subset of Ω can be taken as an event. However, if Ω is uncountable, it may not be possible to take all subsets to be events. For example, if $\Omega = [0, 1]$ and we require that $P(\text{interval}) = \text{length}(\text{interval})$, then it is a well-known example in Lebesgue integration theory that there are subsets of $[0, 1]$ for which P cannot be defined if the three axioms are to be satisfied. However, for what we do in this book, technicalities such as this are not of great importance. We need only be aware of the existence of these problems.

2. Computation of Probabilities

The three axioms of probability make it immediately clear that the probabilities of some events can be computed from those of others. We attempt to develop this idea in this section. This is an important point because in practice it means that we need only start with the probabilities for a subcollection of the events, and compute the rest using the axioms.

Example 2.1 Consider one toss of a single coin. The possible outcomes are "heads" and "tails." There are four possible events: the empty set $\emptyset$, {heads}, {tails}, {heads, tails}. Suppose we know that $P(\{\text{heads}\}) = p$. Then

$$P(\{\text{heads, tails}\}) = 1 \quad \text{by Axiom 1}$$

$$\left.\begin{aligned} P(\{\text{tails}\}) &= 1 - p \\ P(\text{empty set}) &= 0 \end{aligned}\right\} \text{by Axioms 1 and 2}$$

The procedure of starting with the probabilities of a subcollection of events and computing the rest is known as extension. If Ω is finite, we need only the first two axioms for extending a probability. First, let A^c denote the complement of A, i.e., the set of points in Ω that are not in A. Then we must have

(2.1) $$P(A^c) = 1 - P(A)$$

Next, we use the notation A + B to mean A ∪ B when A and B are disjoint, and we use the notation A - B to mean $A \cup B^c$ when B is contained in A. Axiom 2 can be reexpressed as

(2.2) $$P(A + B) = P(A) + P(B)$$

which in turn implies that

(2.3) $$P(A - B) = P(A) - P(B)$$

because B + (A - B) = A.

Using the notation developed above, we can write for two arbitrary sets A and B

(2.4) $$A \cup B = (A - A \cap B) + A \cap B + (B - A \cap B)$$

Figure 1.1 makes this expression obvious.

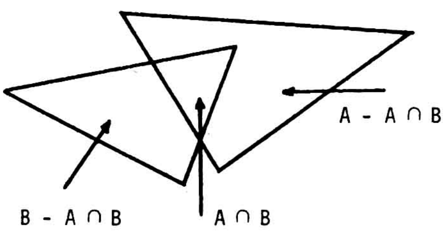

Figure 1.1

Therefore, the probability of A ∪ B is given by

(2.5) $$\begin{aligned} P(A \cup B) &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

Suppose that we start with a collection *S* of events such that *S* is closed under intersection; i.e., whenever A and B are sets in *S*, then A ∩ B is also in *S*. If we know the probability P(A) for every set A in *S*, then by repeated applications of (2.1) and (2.5) we can determine the probability of every set that can be obtained from sets in S by a finite series of unions, intersections, and complementations. The collection of all such sets will be denoted by $B(S)$. A collection of sets, such as $B(S)$, closed under finite boolean set operations is known as an <u>algebra</u>.

<u>Example 2.2</u> Let the experiment be the toss of a die so that $\Omega = \{1, 2, 3, 4, 5, 6\}$. Let $A_0 = \emptyset$ be the empty set and let $A_1, A_2, \cdots, A_6$ be the six sets, each containing just a single number. The collection $S = \{A_0, A_1, A_2, \cdots, A_6\}$ is closed under intersection because the intersection between any two A's is empty. For this case $B(S)$ is the collection of all possible subsets of Ω. There are $2^6 = 64$ such subsets, because each subset is uniquely identified by asking whether it contains the number i for $i = 1, 2, \cdots, 6$. For this problem if we know the probabilities of any five of six sets $A_1, A_2, \cdots, A_6$, we can determine the probability of every subset A of Ω, which is just the sum of the probabilities of the numbers $i = 1, 2, \cdots, 6$ contained in A.

<u>Example 2.3</u> Let $\Omega = [0, 1)$. Let S be the collection of all semiopen intervals of the form $[a, b)$, $0 \le a \le b \le 1$. We take $[a, a)$ to be the empty set. Let the probability of $[a, b)$ be given by

$$P([a, b)) = b - a$$

The collection S is closed under intersection. It can be shown that $B(S)$ is the collection of all sets of the form

$$A = \bigcup_{i=1}^{n} [a_i, b_i)$$

where the intervals $[a_i, b_i)$, $i = 1, 2, \cdots, n$, are disjoint and n can be any finite integer. Clearly, for such an event we have

$$P(A) = P\left(\bigcup_{i=1}^{n} [a_i, b_i)\right) = \sum_{i=1}^{n} (b_i - a_i)$$

The extension of a probability P from S to $B(S)$ makes use of only (2.1) and (2.5), which in turn are derived using only the first two axioms of probability. We now make use of the third axiom, sequential continuity. First, we need to define the concept of convergence for a sequence of sets. By a sequence of sets we mean a countable collection of sets, i.e., a collection of sets $\{A_i, i = 1, 2, \cdots\}$ that can be indexed by the positive integers. Suppose that for every n, $A_{n+1} \supseteq A_n$; then the union $\bigcup_{k=1}^{n} A_k$ must be equal to A_n. Therefore, it is natural to define

$$\lim_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} A_k$$

Similarly, if $A_{n+1} \subseteq A_n$ for every n, then we define

$$\lim_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} A_i$$

For a general sequence $\{A_k, k = 1, 2, \cdots\}$ we set

$$B_n = \bigcup_{k\geq n} A_k$$

$$C_n = \bigcap_{k\geq n} A_k$$

It is always true that $B_{n+1} \subseteq B_n$ and $C_{n+1} \supseteq C_n$ for every n. We say that the sequence of sets $\{A_k\}$ <u>converges</u> if $\lim_{n\to\infty} C_n = \lim_{n\to\infty} B_n$, i.e., if

$$\bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k \tag{2.6}$$

and we take this common limit to be $\lim_{n\to\infty} A_n$. Axiom 3 of probability now reads: If $\{A_k\}$ is a sequence of sets such that (2.6) holds, then

$$\mathcal{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathcal{P}(A_n)$$

<u>Example 2.4</u> For the case considered in Example 2.3, we have

$$[a, b] = \bigcap_{n=1}^{\infty} [a, b + 1/n)$$

Therefore,

$$P([a, b]) = \lim_{n\to\infty} P([a, b + 1/n))$$

$$= \lim_{n\to\infty} (b + 1/n - a) = b - a$$

which shows that [a, b] and [a, b) must have the same probability.

A probability P is a set function, and its natural domain of definition is a collection of sets that is closed under all finite boolean set operations and sequential limits, or equivalently all countable set operations. Such a collection, which we usually denote by A, is known as a <u>σ-algebra</u>. Axiom 3 requires P to be continuous relative to sequential limits. The triplet (Ω, A, P) is known as a <u>probability space</u>.

For a given collection S, there is a smallest σ-algebra that contains S and it is known as the σ-algebra generated by S.

<u>Example 2.5</u> For the S defined in Example 2.3 it is easy to show that every subinterval of [0, 1), closed or open at either end, is in $A(S)$, and P(interval) = length of the interval. For instance, $[a, b] = \lim_{n\to\infty} [a, b + 1/n)$ in the sense of sequential limit of sets. Therefore,

$$P([a, b]) = \lim_{n\to\infty} P(b - a + 1/n) = b - a$$

<u>Example 2.6</u> Consider an experiment with an infinite number of coin tosses. We can take Ω to be the set of all infinite sequences of 1's and 0's, with 1 standing for "heads" and 0 for "tails." Hence each point ω in Ω has the form

$$\omega = (\omega_1, \omega_2, \omega_3, \cdots), \quad \omega_i = 0, 1$$

If we want every subset of Ω to be an event, we can begin with a collection S defined as follows. S contains every set of ω's that is specified by fixing a finite number of ω_i's. For example, "the set of all ω's such that $\omega_1 = 1, \omega_7 = 0, \omega_{16} = 1, \omega_{136} = 0$" is one such set. We can also describe S in another way. Let

$$A_i = \{\omega : \omega_i = 1\}$$

Then, A_i^c is clearly the set $\{\omega : \omega_i = 0\}$. Every set in S is the intersection of a finite number of A_i's and A_i^c's. For example, the set $\{\omega : \omega_1 = 1, \omega_7 = 0, \omega_{16} = 1, \omega_{136} = 0\}$ is $A_1 \cap A_7^c \cap A_{16} \cap A_{136}^c$. If the coin is "fair," then we take for every i

$$P(A_i) = \frac{1}{2} = P(A_i^c)$$

and this is our definition for "fair." If the tosses are "independent," then for every set A in S we take $P(A)$ to be the product of the probabilities of the A_i's and A_j^c's making up A. For example,

$$P(A_1 \cap A_7^c \cap A_{16} \cap A_{136}^c) = P(A_1)P(A_1^c)P(A_{16})P(A_{136}^c)$$

This provides a definition for "independent" tosses. Thus, for independent tosses of a fair coin, for each A in S, $P(A)$ is $1/2^n$, where n is the number of ω_i's that are fixed in A.

3. Independent and Conditional Probability

Let A be an event with $P(A) > 0$. For any event B we can define the conditional probability given A by

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \tag{3.1}$$

The interpretation of $P(B/A)$ is the likelihood that the actual outcome will be in B given the information that the outcome will be in A. We say that two events A and B are independent if

$$P(A \cap B) = P(A)P(B) \tag{3.2}$$

This suggestive terminology comes from the fact that if A and B are independent and $P(A) > 0$, then the conditional probability $P(B/A)$ is just $P(B)$. In other words, given the information that the outcome will be in A does not change the probability of the event B.

A finite collection $\{A_i, i = 1, 2, \cdots, n\}$ of events is said to be an independent collection if every finite subcollection $\{A_{k_1}, A_{k_2}, \cdots, A_{k_m}\}$ has the property

(3.3) $$P\left(\bigcap_{j=1}^{m} A_{k_j}\right) = \prod_{j=1}^{m} P(A_{k_j})$$

An arbitrary collection of events, finite, countable, or uncountable, is said to be independent if every finite subcollection is independent. Events in an independent collection are said to be mutually independent, or simply independent.

Independent sequences (i.e., countable families) of events are of special interest. For example, they lead to a simple result on the interpretation of probability as relative frequency. Suppose that $\{A_1, A_2, \cdots\}$ is an independent sequence of events such that $P(A_i) = p$ is the same for each A_i. Then the number p can be obtained as follows: Once the actual outcome ω of the random experiment is observed, we can determine the number

$$n_N(\omega) = \text{number of } A_i\text{'s containing } \omega \text{ among } A_1, A_2, \cdots, A_N$$

The relative frequency is then given by

(3.4) $$\frac{n_N(\omega)}{N} = \hat{p}_N(\omega)$$

It can be shown (cf. Section 3.2) that for every $\varepsilon > 0$, the event

$$\{\omega : |\hat{p}_N(\omega) - p| \geq \varepsilon\}$$

has a probability less than or equal to $1/4N\varepsilon^2$ which goes to zero as $N \to \infty$. This is one of the simplest versions of the law of large numbers, and can be interpreted as saying that if N is large, then most of the outcomes will yield a relative frequency close to p. For example, given the actual outcome, say ω_0, if we compute the relative frequency $\hat{p}_N(\omega_0)$ for $N = 10^4$, then there is a better than 99% chance that $\hat{p}_N(\omega_0)$ is within 0.5% of the actual probability p. Note, once again, that our premise is always that no more than one outcome is ever observed in a random experiment. The concept of relative frequency is associated with a sequence of events all having the same probability, not with a single event. As in Example 2.6, consider a repeated coin-tossing experiment where an outcome is an infinite sequence of heads and tails. The event A_i is equal to {all outcomes "coming up heads" on the ith toss} for $i = 1, 2, \cdots$. Since $\{A_i\}$ is an independent sequence of events with $P(A_i) = 1/2$ for every i, almost every outcome will have an equal number of heads and tails in the sense of relative frequency. Intuitively, we believe this to be true if the coin is "fair" and the tosses are "independent." Indeed, the assumption that $\{A_i\}$ are independent and equally probable provides a precise definition for "fair coin" and "independent" toss.

Chapter 2. Random Variables

1. Definition and Distribution Function

In practical situations we are usually more interested in a real number that depends on the outcomes rather than on the outcomes themselves. Of course, a real number that depends on the outcomes ω is in fact a real-valued function of ω.

Definition A real random variable is a real-valued function $X(\omega)$, $\omega \in \Omega$, such that for every real number a, the set $\{\omega : X(\omega) < a\}$ is an event.

The requirement that $\{\omega : X(\omega) < a\}$ be an event for every a is imposed in order that we can begin to discuss probability in connection with X.

The probability of $\{\omega : X(\omega) < a\}$ defines a nonnegative-valued function $P_X(a)$, $-\infty < a < \infty$. This function P_X is called the probability distribution function of the random variable X. It has two important properties:

(1) P_X is a nondecreasing function, with $\lim_{a\to\infty} P_X(a) = 1$ and $\lim_{a\to-\infty} P_X(a) = 0$.

(2) P_X is left-continuous, i.e., $\lim_{\varepsilon\downarrow 0} P_X(a - \varepsilon) = P_X(a)$.

We should note that left (rather than right) continuity is in consequence of the strict inequality $X < a$ in the definition of P_X. If we had used $\le$ instead of $<$ in defining P_X, it would be right-continuous. The literature is not standard on this point. It is a matter of taste whether one adopts $<$ or $\le$ in defining P_X. In this book we choose $<$ throughout. The nondecreasing property follows from the fact that a probability is additive and nonnegative, because

$$\begin{aligned} P_X(a + \varepsilon) &= P(\{\omega : X(\omega) < a + \varepsilon\}) \\ &= P(\{\omega : X(\omega) < a\} + \{\omega : a \le X(\omega) < a + \varepsilon\}) \\ &= P_X(a) + P(\{\omega : a \le X(\omega) < a + \varepsilon\}) \end{aligned}$$

The left-continuity follows from sequential continuity of probability, because

$$P_X(a) - P_X(a - \tfrac{1}{n}) = P(\{\omega : a - \tfrac{1}{n} \le X(\omega) < a\} \xrightarrow[n \to \infty]{} 0$$

In a very real sense any probabilistic question concerning X can be answered directly once we know its probability distribution function P_X. The distribution function P_X, being a real-valued function of a real variable, is much simpler to deal with than $\mathcal{P}$, which is a function defined on sets. In practice, the situation is made even simpler by the fact that P_X is frequently one of two forms:

(1) P_X is constant except for jumps at a discrete sequence of points $x_1, x_2, \cdots$.

(2) P_X is of the form

$$P_X(a) = \int_{-\infty}^{a} p_X(x)\, dx$$

In the first case we say that P_X is a discrete distribution and interpret the situation to mean that X can only take on the values $x_1, x_2, \cdots$ with nonzero probability. In the second case P_X is said to be absolutely continuous and the integrand p_X is called the probability density function. If p_X is continuous at a, then of course we have

$$p_X(a) = \frac{d}{da} P_X(a)$$

Probability density functions are nonnegative and $\int_{-\infty}^{\infty} p_X(x)\, dx$ is always equal to 1. Although p_X is not a probability, we can interpret $p_X(x)\, dx$ to be the probability of the event $\{\omega : x < X(\omega) < x + dx\}$, so that $p(x)$ is probability per unit interval, hence the name "density."

Example 1.1 Consider the repeated coin-tossing experiment described in Example 1.2.6. Let $X(\omega)$ be the number of 1's among the first N components of ω. X is a random variable taking values 0, 1, 2, $\cdots$, N. It can be shown that

$$\mathcal{P}(\{\omega : X(\omega) = k\}) = \frac{1}{2^N} \binom{N}{k}$$

$$= \frac{1}{2^N} \frac{N!}{k!(N-k)!}$$

Example 1.2 Let $\Omega = [0, 1)$ and

$$\mathcal{P}(\text{interval}) = \text{length of interval}$$

as described in Example 1.2.5. Let $X(\omega) = \omega^2$, $0 \le \omega < 1$. It is clear that X