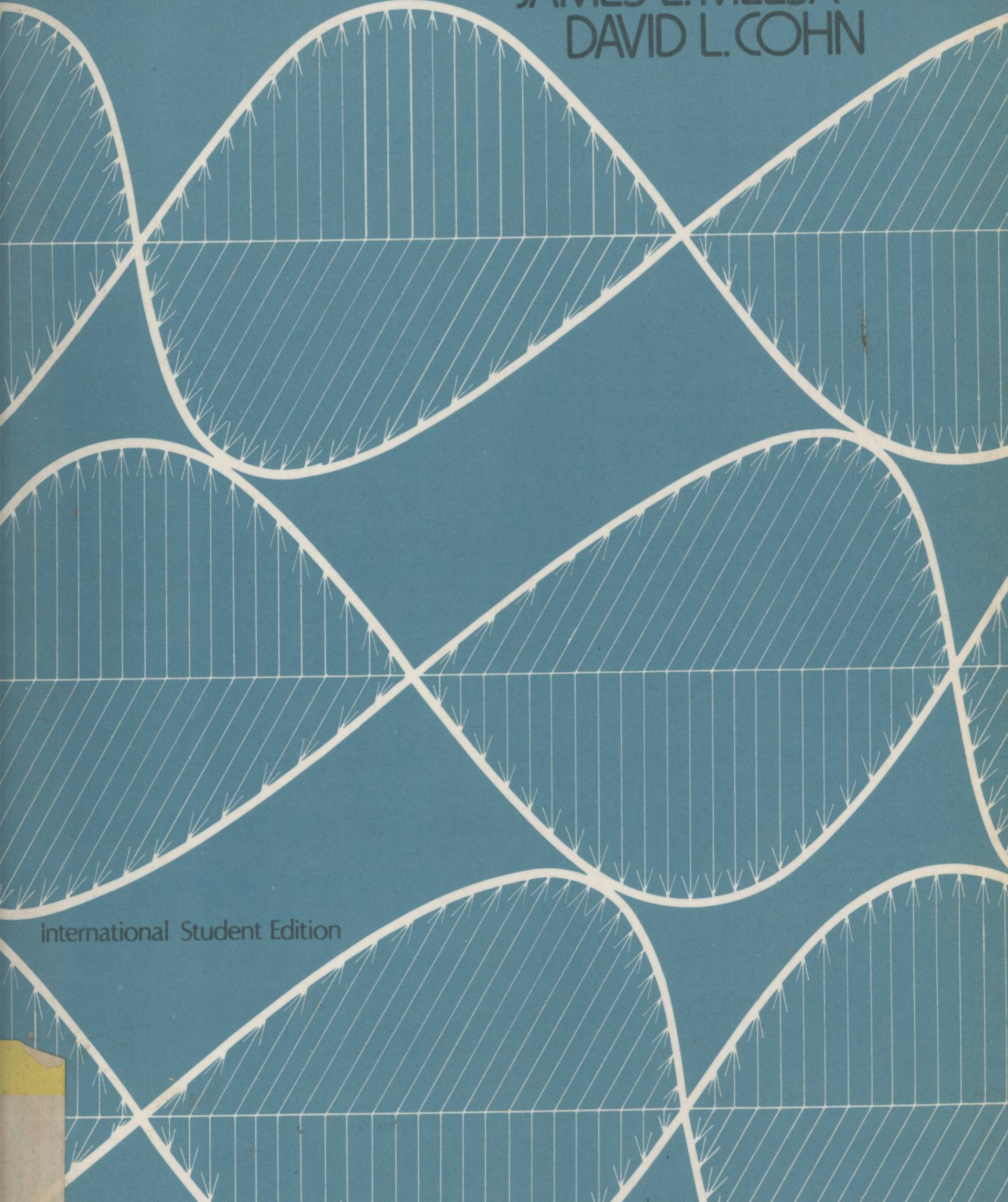
DECISION AND ESTIMATION THEORY

JAMES L. MELSA
DAVID L. COHN



International Student Edition

DECISION AND ESTIMATION THEORY

James L. Melsa
David L. Cohn

Department of Electrical Engineering
University of Notre Dame

INTERNATIONAL STUDENT EDITION

McGRAW-HILL KOGAKUSHA, LTD.

Tokyo Auckland Bogota Guatemala Hamburg Johannesburg Lisbon London Madrid Mexico New Delhi Panama Paris San Juan São Paulo Singapore Sydney

DECISION AND ESTIMATION THEORY

INTERNATIONAL STUDENT EDITION

This book was set in Times Roman by A Graphic Method Inc.
The editors were Frank Cerra and Frances A. Neal;
the cover was designed by Rafael Hernandez;
the production supervisor was Dominick Petrellese.
The drawings were done by J & R Services, Inc.

Library of Congress Cataloging in Publication Data

Melsa, James L
Decision and estimation theory.

Includes bibliographical references and index.
1. Statistical decision. 2. Estimation theory.
I. Cohn, David L., date joint author. II. Title.
QA279.4.M44 519.5'4 77-22759
ISBN 0-07-041468-8

TOSHO PRINTING CO., LTD., TOKYO, JAPAN

DECISION AND ESTIMATION THEORY

TO THOSE WHO FIRST TAUGHT US ABOUT DECISIONS

Ann and Louis Melsa
Marjorie and Nathan Cohn

PREFACE

Decision and estimation theory are basic tools which are used in many areas of communications, control theory, and system theory. This book provides a unified presentation of these important tools. The step-by-step development is designed to provide the reader with a solid background in this complex area.

The book is intended as a textbook for a one-semester course in decision and estimation theory. It is designed for advanced seniors or first-year graduate students. We assume that the reader has had prior exposure to probability and random process theory. This background may have been acquired from a course using one of the many available books such as "Probability, Random Variables, and Stochastic Processes" by A. Papoulis, "An Introduction to Probability and Stochastic Processes" by J. L. Melsa and A. P. Sage, or "Probability and Random Processes" by W. B. Davenport. Alternatively, the student may have been exposed to these concepts elsewhere. The major elements of probability theory and stochastic processes that are required in this text are summarized in Chapter 2. A student who can work the exercises at the end of Chapter 2 probably has sufficient background in this area.

Decision and estimation problems have essentially the same structure. Some sort of *message* is generated at a source which causes an *observation* at a receiver. The message and observation are only stochastically related. The objective is to determine a rule which forms a "best guess" of the message based on the observation.

This book deals with the problem of making decisions about both discrete and continuous messages. Although there are important differences between the discrete and the continuous problems, they are treated in a parallel fashion. Chapters 1 to 7 deal with the discrete decision problems, while Chapters 8 to 11 are concerned with continuous estimation problems.

Chapter 3 begins the treatment of decision problems with the simplest possible problem structure. It considers the case of binary decisions and a single observation. The discussion begins with the fewest assumptions concerning the system model. Subsequent sections add structure and analyze the results of these additions. In Chapter 4, the problem is generalized to include multiple and waveform observations.

The problem is generalized still further in Chapter 5 where multiple decision problems are considered. This chapter ends with the erasure decision problem, which is used as an introduction to the sequential decision problem in Chapter 6. Chapter 6 includes a careful motivation of the Wald test and is one of the major features of the book. The discussion of decision problems is concluded in Chapter 7, which summarizes the composite decision problem and presents some simple nonparametric decision methods.

Chapters 8 to 11 examine the impact of enlarging the message space from a finite number of elements to an uncountably infinite number of elements. Chapter 8 discusses simple estimation structures and parallels the development in Chapter 3. The important case of estimation with gaussian noise is discussed in Chapter 9. In particular, sequential estimation and nonlinear estimation are described. Chapter 10 examines some of the properties of estimators. Finally, the important state estimation problem and the Kalman filter are introduced in Chapter 11.

No work of this nature is prepared in a vacuum. Therefore, we would like to acknowledge those who have established the atmosphere which enabled us to undertake and complete this project. Deans Thomas Martin at Southern Methodist University and Joseph Hogan at University of Notre Dame deserve special mention for the leadership they provided. Our colleagues and students at these two universities, who suffered through the earlier stages with us, also merit recognition. A succession of secretaries typed and retyped the manuscript. And, finally, we acknowledge the readers; they may find, as we did, that the most intensive learning takes place while correcting the mistakes of others.

James L. Melsa
David L. Cohn

DECISION AND ESTIMATION THEORY

CONTENTS

CHAPTER
ONE

INTRODUCTION

1.1 STRUCTURE OF DECISION AND ESTIMATION PROBLEMS

All the decision and estimation problems discussed in this book have the same basic structure. This structure involves four steps: first, something happens; second, this happening is relayed to an observer by some form of signaling mechanism; third, a noisy observation of this signal is made; and finally, the observer must reach a decision concerning the happening. This problem structure is shown schematically in Fig. 1.1-1. We are involved with problems of this sort all the time. Consider, for example, the medical doctor attempting to diagnose a patient's illness. The happening in this case is the illness. The signal is the set of medical tests which the doctor can make, each of which is corrupted by various noise sources including physical abnormalities the patient may have. The doctor then must decide what illness the patient has.

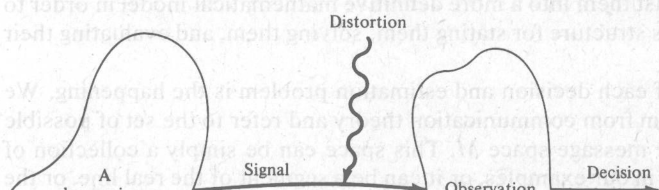


Figure 1.1-1 Structure of decision and estimation problems.

Another example of this type of problem is encountered in modeling a communication system. Suppose that a message is to be sent to a remote location. For the first step, a sender selects a message from a predetermined set of messages. For example, the sender might decide which key on a teletypewriter keyboard to push. The signal corresponding to this message is then transmitted. For example, pushing the key might generate a sequence of pulses that are transmitted over a telephone line. The transmitted signal is distorted by noise in the transmission media so that the signal observed at the receiver does not match the transmitted signal. The receiver must then use the distorted signal to make a best guess of which key was pushed. An appropriate definition of "best guess" here might be to select the key with the highest probability of having been pushed. Alternatively, best guess might mean to select the key which minimizes the probability of error.

A slightly different example arises in a biomedical experiment. An experimentalist may wish to determine if a particular optic nerve in a frog's eyes responds to a light stimulus. In this case, the happening occurs when the experimentalist selects one of the many thousands of nerve fibers. The observations are the electrical measurements made on the nerve when a light stimulus is applied to the frog's eye. In this example, the goal is not to guess which nerve was selected. Rather, there are only three possible decisions: the fiber responds; it does not respond; or more information is needed. If the third decision is reached, the experiment is repeated. The information for the first trial is available in formulating the decision rule for the next trial.

Note the importance of the distortion which occurs to the signal before the observation is made. Without this distortion, the signal would be directly observed and there would be no uncertainty concerning what happened. It is this distortion that makes decision and estimation theory important.

There are many other practical problems that fit this general problem structure, and the methods for formulating and solving these problems are the topic of this book. Although we treat such problems intuitively all the time, it is important that we cast them into a more definitive mathematical model in order to develop a rigorous structure for stating them, solving them, and evaluating their solution.

At the root of each decision and estimation problem is the happening. We shall borrow a term from communication theory and refer to the set of possible happenings as the message space M. This space can be simply a collection of discrete points as in our examples, or it can be a segment of the real line, or the points in an N-dimensional euclidean space, or even a set of time functions. Often, but not always, there may be a probability measure defined on this message space that tells us the way in which the messages are selected.

Generally, if M has a finite number of points, the problem is called a *decision* or *detection problem*. If the number of points are uncountably infinite, the problem is referred to as an *estimation problem*. For the finite case, we say that there are K points in the space and define them as m_k where k takes on the values from 1 to K. For the estimation problem, messages will be generalized as a continuous random vector $\mathbf{m}$ with dimension K. If the messages are time functions, we shall use the random process $m(t)$.

The concept of a signal space S is used to isolate the portion of the problem where information is generated from the portion where that information is transmitted. In the communication example this distinction was clear. The message was a letter on the key, and the signal was a series of pulses. In the biomedical case, the message was a particular nerve selected and the signal was a reaction of that nerve to the light stimulus. In general, we shall use a lower case s to identify points in the signal space. As with messages, s could be either a discrete or continuous random variable, a random vector, or a stochastic process. The distinction will be made clear as needed for a particular problem. In general, the signal statistical type is not necessarily the same as that of the message. However, on occasion the message and the signal will be identical. We assume that there is a unique and invertible mapping between elements of the message space and elements of the signal space. That is, each message generates a uniquely defined signal, and each signal is generated by only one message.

The third part of our mathematical model is a set of possible observations known as the *observation space* Z. Frequently, we shall be trying to model some physical mechanism that relates the signal and the observation. When this happens, S and Z will be the same statistical type. However, there may be times when there is no physical relationship between S and Z, only a statistical one. In any case, we shall assume that for each point S, there is a conditional probability measure defined on Z. For example, when Z consists of a segment of the real line, the probability density function $p(z|s)$ will be defined for each z in the observation space Z and each s in the signal space S. Often one says that the signals are mapped into the observation space by means of *probabilistic transition mechanism*, or, in the communication sense, this is the *channel model*.

Finally, a decision must be reached. There is as much flexibility in the size of the decision space D as there is in the message space. In many problems, however, D and M actually consist of the same set of points. In this case, the decision is usually a best guess of what the message was. This was the case in our second example but not the third. The relationship between the observation space and the decision space is called the *decision rule*. In a typical decision or estimation problem, the message, signal, observation, and decision spaces are all defined and the relationships between M, S, and Z are specified. The missing piece is the decision or estimation rule, that is, the mapping from the observation to a decision. Therefore, in this book we shall be concerned primarily with designing decision and estimation rules. Typically these rules will be deterministic; in other words, for each observation there will be a unique decision. These four spaces and their interrelations are shown schematically in Fig. 1.1-2.

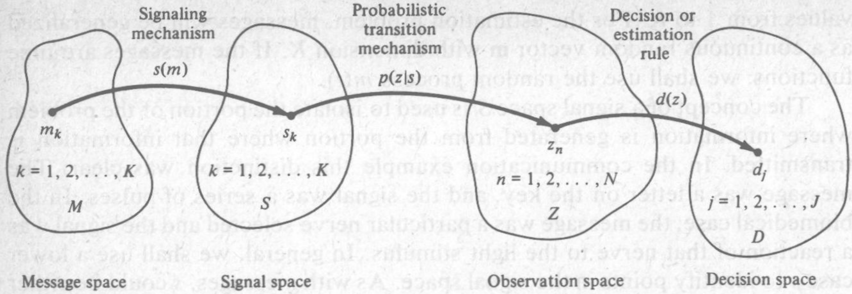


Figure 1.1-2 Decision and estimation problems.

1.2 OUTLINE OF THE BOOK

Following the introductory material of this chapter, Chap. 2 presents a brief review of certain concepts and notations in probability theory that will be used throughout the developments in this book. Following this review in Chap. 2, the remainder of the book can be divided into two broad classes: Chaps. 3 to 7 are directed to solutions of decision problems, while Chaps. 8 to 11 concern estimation problems.

We begin our study of decision theory problems in Chap. 3 with simple binary decision problems having only a single scalar observation. With this class of problems we can introduce most of the concepts within the simplest possible mathematical framework. In the binary decision problem, the message and decision spaces contain only two elements. Several different formulations and solutions of this problem are considered.

In Chap. 4, we will broaden the class of binary decision problems to include those where the observation is more complicated than the single scalar observation. Initially we let the observation be a set of scalar variables, and then we consider time-waveform observations. In particular, we will use the gaussian problem as a focal point for the development of several interesting concepts involving the signal-space approach to decision problems.

Chapter 5 considers another extension of the simple binary decision problem treated in Chap. 3. Here we consider problems for which the decision space has more than two elements. We begin with the general bayesian approach and then develop a more practically meaningful result using the probability-of-error criterion. Finally, we consider erasure decision problems.

In all the decision problems considered in these early chapters, it is assumed that the number of observations is fixed. Sometimes, however, it is possible for the receiver to take additional observations before making a final decision. Such problems are called *sequential decision problems* and are the topic of Chap. 6. We begin with treatment of the bayesian approach to this problem and later consider the Wald sequential-decision-theory approach.

In Chap. 7, we examine a number of methods for handling decision problems when the conditional probability density of the observation is not completely known. We begin with the case where the conditional density is known in its basic form but where there are certain unknown parameters. This type of problem is known as a *composite decision problem*. In the second part of Chap. 7, we consider another class of problems, known as *nonparametric decision problems*, in which the conditional probability density function is assumed to be unknown, except for general properties such as symmetry.

Chapter 8 begins our discussion of estimation problems. Because estimation problems are really an extension of multiple decision problems treated in Chap. 5, it is possible to make use of a great amount of the theory presented in the preceding chapters in the development of estimation methods. Chapter 8 considers the development of several different classes of estimators and shows their general properties and some examples of their use.

In Chap. 9, the general problem of estimating a gaussian vector in the presence of gaussian noise is examined. It is a problem of considerable interest both because it has many applications and because a complete solution is easy to find.

Chapter 10 is directed to the study of the properties of different forms of estimators. One of the major differences between the estimation problem and the decision problem is the difficulty in describing how good a given estimation algorithm is. Chapter 10 attempts to quantify certain desirable properties of estimators and show how they can be developed and used.

The state estimation problem is considered in Chap. 11. In Chaps. 8 to 9, we estimated a constant parameter. Now we wish to estimate the time-varying state of a dynamic system. In particular, our interest is in a class of linear unbiased minimum-error-variance sequential-state estimators referred to as the *Wiener-Kalman filters*.

CHAPTER
TWO

REVIEW OF PROBABILITY THEORY

In this chapter, some of the basic notions of probability and random processes will be reviewed. Further details regarding these concepts can be found in any one of several textbooks on probability and stochastic processes, including Melsa and Sage (1973) and Papoulis (1965). The objectives are to refresh the reader's memory and present notation that will be used throughout the book. Whenever possible, the examples will be those that occur most often in decision and estimation. The chapter is divided into three sections: Section 2.1 deals with simple probability theory; Sec. 2.2 introduces random variables; and Sec. 2.3 treats stochastic processes.

In any presentation of probability theory, a choice must be made between exactness and economy of notation. In an introductory presentation of the subject, it is generally best to use the most exact notation possible. The same is also true of advanced presentations which emphasize the mathematical aspects of the topic. In this text, however, probability theory is to be used as a tool to solve a set of interesting problems. Therefore, whenever possible, simple notational forms will be used. One objective of this chapter is to present this simplified notation.

2.1 DISCRETE PROBABILITY THEORY

Probability theory is an attempt to model mathematically a set of real-world situations. A typical situation to be modeled is one where an experiment is performed and the result of the experiment is observed. The probabilistic model, or probability space, consists of three components: a set of possible alternatives, or *sample points*; a set of *events* defined on the sample points; and a *probability measure* defined on the events. The set of sample points, or sample space Ω, can be visualized as the set of all possible outcomes of the experiment.