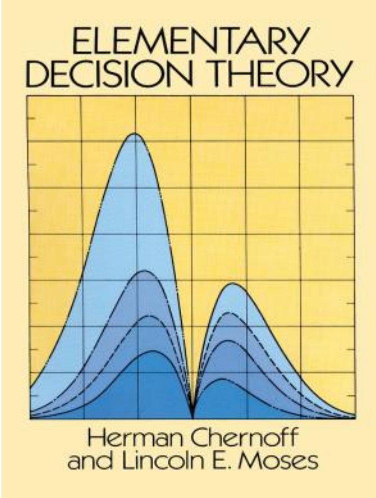
ELEMENTARY
DECISION THEORY
Herman Chernoff
and Lincoln E. Moses

Elementary Decision Theory

HERMAN CHERNOFF
Professor of Statistics, Harvard University

LINCOLN E. MOSES
Professor of Statistics, Stanford University

Dover Publications, Inc.
NEW YORK

This Dover edition, first published in 1986, is an unabridged, corrected republication of the work first published by John Wiley & Sons, Inc., New York, in 1959.

Library of Congress Cataloging-in-Publication Data

Chernoff, Herman.

Elementary decision theory.

Reprint. Originally published: New York: Wiley, 1959. (Wiley publications in statistics)

Includes index.

1. Mathematical statistics. 2. Statistical decision. I. Moses, Lincoln E. II. Title.

QA279.4.C47 1986 519.5 86-16842

9780486143774

Manufactured in the United States by Courier Corporation
65218106
www.doverpublications.com

Dedicated to the memory
of M. A. Girshick

Preface

In recent years, Statistics has been formulated as the science of decision making under uncertainty. This formulation represents the culmination of many years of development and, for the first time, furnishes a simple and straightforward method of exhibiting the fundamental aspects of a statistical problem. Earlier representations had serious gaps which led to confusion and misunderstanding, especially on the part of elementary students without well-developed statistical intuition.

This book is the result of nine years of experience at Stanford in teaching a first course in Statistics from the decision making point of view. The course was first started by the late M. A. Girshick and this book may be regarded in part as an extension of his teaching.

A “first course” is contained in the first seven chapters. Our experience has been mainly with social science students in a five-unit, one-quarter course. Here we covered the seven chapters in detail, including the optional sections marked (†), and portions of the other chapters.

A background of high school mathematics suffices for this course, and considerable effort has been taken to by-pass involved computational reasonings which confuse the inexperienced. On the other hand, there has been no reluctance to use symbols, and the militantly non-mathematical student will not enjoy this book.

The mathematical novice should find in this book a well-motivated introduction to certain important and uncomplicated mathematical notions such as set, function, and convexity. Students who have a strong background in mathematics will find it profitable to spend time on the Appendixes which have the proofs of basic results in decision theory.

The reader will observe that new topics and ideas are introduced by examples. Also, certain exercises carry an essential burden in the development of the material. These exercises are starred and should be assigned. The teacher should take care to note that a few of the exercises in the first three chapters and in Chapter 7 are liable to be time-consuming.

In teaching the course, we have found it expedient to cover the material in Chapter 1 in one lecture and to spend another hour or two on the associated homework and classroom discussion. Chapter 2 was customarily disposed of rapidly, in about three hours. Considerable time was devoted to Chapter 5 which presents in detail most of the underlying decision theory.

The last three chapters may be studied in any order. Chapter 8 consists of a relatively informal discussion of model building. Chapters 9 and 10 treat classical statistical theory from the decision theory point of view and are rather technical. Ordinarily we have covered only portions of them. We regard it as inadvisable to attempt to do much in these chapters unless the students have some previous background in statistics or mathematics.

This book does not have a treatment of classical statistical methodology. It is our hope to follow this volume with a second one presenting existing methodology in the decision theory framework. Thus, this book will not be of much use to students who would like a single course devoted to the study of a few generally applicable statistical methods. Rather, we feel, that this book is well designed for those who are interested in the fundamental ideas underlying statistics and scientific method and for those who feel that they will have enough need for Statistics to warrant taking more than one course.

HERMAN CHERNOFF
LINCOLN E. MOSES

Stanford University
March 1959

Acknowledgments

Our greatest debt is to Sir Ronald A. Fisher, Jerzy Neyman, Egon S. Pearson, John von Neumann, and Abraham Wald. The present state of statistical theory is largely the result of their contributions which permeate this book. Particularly great is our debt to the late Abraham Wald who formulated statistics as decision making under uncertainty.

We acknowledge with warm thanks the keen and constructive comments of William H. Kruskal; we have acted upon many of his suggestions and have improved the book thereby. We thank Patrick Suppes for helpful comments on certain problems of mathematical notation. Harvey Wagner suggested that we incorporate the outline of a proof of the existence of the utility function as a handy reference for the mathematically sophisticated reader. He also proposed a proof which we found useful. Max Woods was especially effective in his careful preparation of the answers to the exercises and his detailed reading of the manuscript and proofs. Others who helped in reading the manuscript and proofs were Arthur Albert, Stuart Bessler, Judith Chernoff, Mary L. Epling, Joseph Kullback, Gerald Lenthall, Edward B. Perrin, Edythalena Tompkins, and Donald M. Ylvisaker. The completion of this manuscript would have long been delayed except for the unstinting cooperation of Carolyn Young, Charlotte Austin, Sharon Steck, and Max Woods. We thank our wives Judith Chernoff and Jean Moses for their support and encouragement.

The authors wish to express their appreciation to E. S. Pearson, C. Clopper, P. C. Mahalanobis, F. E. Croxton, and C. M. Thompson for the use of their tables. We are indebted to Professor Sir Ronald A. Fisher, Cambridge, to Dr. Frank Yates, Rothamsted, and to Messrs. Oliver and Boyd Ltd., Edinburgh, for permission to reprint Tables III, IV, and XXXIII from their book, Statistical Tables for Biological, Agricultural, and Medical Research.

H. C.
L. E. M.

Table of Contents

CHAPTER 1

Introduction

1. INTRODUCTION

Beginning students are generally interested in what constitutes the subject matter of the theory of statistics. Years ago a statistician might have claimed that statistics deals with the processing of data. As a result of relatively recent formulations of statistical theory, today's statistician will be more likely to say that statistics is concerned with decision making in the face of uncertainty. Its applicability ranges from almost all inductive sciences to many situations that people face in everyday life when it is not perfectly obvious what they should do.

What constitutes uncertainty? There are two kinds of uncertainty. One is that due to *randomness.* When someone tosses an ordinary coin, the outcome is random and not at all certain. It is as likely to be heads as tails. This type of uncertainty is in principle relatively simple to treat. For example, if someone were offered two dollars if the coin falls heads, on the condition that he pay one dollar otherwise, he would be inclined to accept the offer since he "knows" that heads is as likely to fall as tails. His knowledge concerns the *laws of randomness* involved in this particular problem.

The other type of uncertainty arises when it is not known which laws of randomness apply. For example, suppose that the above offer were made in connection with a coin that was obviously bent. Then one could assume that heads and tails were not equally likely but that one face was probably favored. In statistical terminology we shall equate the laws of randomness which apply with the *state of nature.*

What can be done in the case where the state of nature is unknown? The statistician can perform relevant experiments and take observations. In the above problem, a statistician would (if he were permitted) toss the coin many times to estimate what is the state of nature. The decision on whether or not to accept the offer would be based on his estimate of the state of nature.

One may ask what constitutes enough observations. That is, how many times should one toss the coin before deciding? A precise answer would be difficult to give at this point. For the time being it suffices to say that the answer would depend on (1) the cost of tossing the coin, and (2) the cost of making the wrong decision. For example, if one were charged a nickel per toss, one would be inclined to take very few observations compared with the case when one were charged one cent per toss. On the other hand, if the wager were changed to $2000 against $1000, then it would pay to take many observations so that one could be quite sure that the estimate of the state of nature were good enough to make it almost certain that the right action is taken.

It is important to realize that no matter how many times the coin is tossed, one may never know for sure what the state of nature is. For example, it is possible, although very unlikely, that an ordinary coin will give 100 heads in a row. It is also possible that a coin which in the long run favors heads will give more tails than heads in 100 tosses. To evaluate the chances of being led astray by such phenomena, the statistician must apply the theory of probability.

Originally we stated that statistics is the theory of decision making in the face of uncertainty. One may argue that, in the above example, the statistician merely estimated the state of nature and made his decision accordingly, and hence, decision making is an overly pretentious name for merely estimating the state of nature. But even in this example, the statistician does more than estimate the state of nature and act accordingly. In the $2000 to $1000 bet he should decide, among other things, whether his estimate is good enough to warrant accepting or rejecting the wager or whether he should take more observations to get a better estimate. An estimate which would be satisfactory for the $2 to $1 bet may be unsatisfactory for deciding the $2000 to $1000 bet.

2. AN EXAMPLE

To illustrate statistical theory and the main factors that enter into decision making, we shall treat a simplified problem in some detail. It is characteristic of many statistical applications that, although real problems are too complex, they can be simplified without changing their essential characteristics. However, the applied statistician must try to keep in mind all assumptions which are not strictly realistic but are introduced for the sake of simplicity. He must do so to avoid assumptions that lead to unrealistic answers.

Example 1.1. The Contractor Example. Suppose that an electrical contractor for a house knows from previous experience in many communities that houses are occupied by only three types of families: those whose peak loads of current used are 15 amperes (amp) at one time in a circuit, those whose peak loads are 20 amp, and those whose peak loads are 30 amp. He can install 15-amp wire, or 20-amp wire, or 30-amp wire. He could save on the cost of his materials in wiring a house if he knew the actual needs of the occupants of that house. However, this is not known to him.

One very easy solution to the problem would be to install 30-amp wire in all houses, but in this case he would be spending more to wire a house than would actually be necessary if it were occupied by a family who used no more than 15 amp or by one that used no more than 20 amp. On the other hand, he could install 15-amp wire in every house. This solution also would not be very good because families who used 20 or 30 amp would frequently burn out the fuses, and not only would he have to replace the wire with more suitable wire but he might also suffer damage to his reputation as a wiring contractor.

TABLE 1.1

LOSSES INCURRED BY CONTRACTOR

Actions / States of Nature	a_1 Install 15 amp	a_2 Install 20 amp	a_3 Install 30 amp
θ_1—family uses 15 amp	1	2	3
θ_2—family uses 20 amp	5	2	3
θ_3—family uses 30 amp	7	6	3

Table 1.1 presents a tabulation of the losses which he sustains from taking various actions for the various types of users.

The thetas (θ) are the possible categories that the occupants of a particular house fall into; or they are the possible states of nature. These are: θ_1—the family has peak loads of 15 amp; θ_2—the family has peak loads of 20 amp; and θ_3—the family has peak loads of 30 amp.[1]

The α's across the top are the actions or the different types of installations he could make. The numbers appearing in the table are his own estimates of the loss that he would incur if he took a particular action in the presence of a particular state.

For example, the 1 in the first row represents the cost of the 15-amp wire. The 2 in the first row represents the cost of the 20-amp wire, which is more expensive since it is thicker.[2]

In the second row we find a 5 opposite state θ_2, under action α_1. This reflects the loss to the contractor of installing 15-amp wire in a home with 20-amp peak loads; cost of reinstallation, and damage to his reputation, all enter into this number. It is the result of a subjective determination on his part; for one of his competitors this number might be, instead, a 6. Other entries in the table have similar interpretations.

Since he could cut down the losses incurred in wiring a house if he knew the value of θ for the house (i.e., what were the electricity requirements of the occupant), he tries to learn this by performing an experiment. His experiment consists of going to the future occupant and asking how many amperes he uses. The response is always one of four numbers: 10, 12, 15, or 20. From previous experience it is known that families of type θ_1, (15-amp users) answer z_1, (10 amp) half of the time and z_2 (12 amp) half of the time; families of type θ_2 (20-amp users) answer z_2 (12 amp) half of the time and z_3, (15 amp) half of the time; and families of type θ_3 (30-amp users) answer z_3, (15 amp) one-third of the time and z_4 (20 amp) two-thirds of the time. These values are shown in Table 1.2. In fact, the entries represent the probabilities of observing the z values for the given states of nature.

TABLE 1.2

FREQUENCY OF RESPONSES FOR VARIOUS STATES OF NATURE IN THE CONTRACTOR EXAMPLE

Observations / States of Nature	z_1 (10 amp)	z_2 (12 amp)	z_3 (15 amp)	z_4 (20 amp)
θ_1	1/2	1/2	0	0
θ_2	0	1/2	1/2	0
θ_3	0	0	1/3	2/3

The contractor now formulates a strategy (rule for decision making) which will tell him what action to take for each kind of observation. For instance, one possible rule would be to install 20-amp wire if he observes z_1; 15-amp wire if he observes z_2; 20-amp wire if he observes z_3; and 30-amp wire if he observes z_4. This we symbolize by s=(α_2, α_1 α_2, α_3), where the first α_2 is the action taken if our survey yields z_1; α_1 is the action taken if z_2 is observed; the second α_2 corresponds to z_3; and α_3 corresponds to z_4.

Table 1.3 shows five of the 81 possible strategies that might be employed, using the above notation.

TABLE 1.3

STRATEGIES (RULES FOR DECISION MAKING)

Strategies \ Observations	z_1 (10 amp)	z_2 (12 amp)	z_3 (15 amp)	z_4 (20 amp)
s_1	α_1	α_1	α_2	α_3
s_2	α_1	α_2	α_3	α_3
s_3	α_3	α_3	α_3	α_3
s_4	α_1	α_1	α_1	α_1
s_5	α_3	α_3	α_2	α_1

Note that s_2 is somewhat more conservative than s_1. Both s_3 and s_4, completely ignore the data. The strategy s_5 seems to be one which only a contractor hopelessly in love could select.

How shall we decide which of the various strategies to apply?

First, we compute the average loss that the contractor would incur for each of the three states and each strategy. For the five strategies, these losses are listed in Table 1.4.

TABLE 1.4

AVERAGE LOSS IN CONTRACTOR EXAMPLE

States of Nature \ Strategies	s_1	s_2	s_3	s_4	s_5
θ_1—family uses 15 amp	1	1.5	3	1	3
θ_2—family uses 20 amp	3.5	2.5	3	5	2.5
θ_3—family uses 30 amp	4	3	3	7	6.67

They are computed in the following fashion:

First we compute the action probabilities for s_1, = (α_1, α_1 α_2, α_3). If θ_1 is the state of nature, we observe z_1 half the time and z_2 half the time (see Table 1.2). If s_1 is applied, action α, is taken in either case, and actions α_2 and α_3 are not taken. If θ_2 is the state of nature, we observe z_2, half the time and z_3 half the time. Under strategy s_1, this leads to action α_1 with probability 1/2, action α_2 with probability 1/2, and action α_3 never. Similarly, under θ_3, we shall take action α_1, never, α_2 with probability 1/3, and α_3 with probability 2/3. These results are summarized in the *action probabilities* for s_1 (Table 1.5) which are placed next to the losses (copied from Table 1.1).

If θ_1 is the state of nature, action α_1 is taken all of the time, giving a loss of 1 all of the time. If θ_2 is the state of nature, action α_1 yielding a loss of 5 is taken half the time and action α_2 yielding a loss of 2 is taken half the time. This leads to an average loss of

$$5 \times 1/2 + 2 \times 1/2 = 3.5.$$

Similarly the average loss under θ_3 is

$$6 \times 1/3 + 3 \times 2/3 = 4.$$

Thus the column of *average losses* corresponding to s_1 has been computed. The corresponding tables for strategy s_2 are indicated in Table 1.5. The other strategies are evaluated similarly.

In relatively simple problems such as this one, it is possible to compute the average losses with less writing by juggling Tables 1.1, 1.2, and 1.3 simultaneously.

Is it clear now which of these strategies should be used? If we look at the chart of average losses (Table 1.4), we see that some of the strategies give greater losses than others. For example, if we compare s_5 with s_2, we see that in each of the three states the average loss associated with s_5 is equal to or greater than that corresponding to s_2. The contractor would therefore do better to use strategy s_2 than strategy s_5 since his average losses would be less for states θ_1 and θ_3 and no more for θ_2. In this case, we say " s_2 dominates s_5." Likewise, if we compare s_4 and s_1, we see that except for state θ_1 where they were equal, the average losses incurred by using s_4 are larger than those incurred by using s_1. Again we would say that s_4 is dominated by strategy s_1. It would be senseless to keep any strategy which is dominated by some other strategy. We can thus discard strategies s_4 and s_5. We can also discard s_3 for we find that it is dominated by s_2.

TABLE 1.5

LOSSES, ACTION PROBABILITIES, AVERAGE LOSS

States of Nature	Losses			Action Probabilities			Average Loss
				For $s_1 = (a_1, a_1, a_2, a_3)$			
	a_1	a_2	a_3	a_1	a_2	a_3	
θ_1	1	2	3	1	0	0	1
θ_2	5	2	3	1/2	1/2	0	3.5
θ_3	7	6	3	0	1/3	2/3	4
				For $s_2 = (a_1, a_2, a_3, a_3)$			
	a_1	a_2	a_3	a_1	a_2	a_3	
θ_1	1	2	3	1/2	1/2	0	1.5
θ_2	5	2	3	0	1/2	1/2	2.5
θ_3	7	6	3	0	0	1	3

If we were to confine ourselves to selecting one of the five listed strategies, we would need now only choose between s_1 and s_2. How can we choose between them? The contractor could make this choice if he had a knowledge of the percentages of families in the community corresponding to states θ_1, θ_2, and θ_3. For instance, if all three states are equally likely, i.e., in the community one-third of the families are in state θ_1, one-third in state θ_2, and one-third in state θ_3, then he would use s_2, because for s_2 his average loss would on the average be

$$1.5 \times 1/3 + 2.5 \times 1/3 + 3 \times 1/3 = 2.33$$

whereas, for s, his average loss would on the average be

$$1 \times 1/3 + 3.5 \text{ x } 1/3 + 4 \text{ x } 1/3 = 2.83.$$

However, if one knew that in this community 90% of the families were in state θ_1 and 10% in θ_2, one would have the *average losses* of

$$\begin{aligned} 1 \times 0.9 + 3.5 \times 0.1 &= 1.25 \qquad \text{for } s_1 \\ 1.5 \times 0.9 + 2.5 \times 0.1 &= 1.60 \qquad \text{for } s_2 \end{aligned}$$

and s_1 would be selected. Therefore, the strategy that should be picked depends on the relative frequencies of families in the three states. Thus, when the actual proportions of the families in the three classes are known, a good strategy is easily selected. *In the absence of such knowledge, choice is inherently difficult*. One principle which has been *suggested* for choosing a strategy is called the “minimax average loss rule.” This says, “Pick that strategy for which the largest average loss is as small as possible, i.e., minimize the maximum average loss.” Referring to Table 1.4, we see that, for s_1, the maximum average loss is 4 and for s_2 it is three.

The minimax rule would select s_2. This is clearly a pessimistic approach since the choice is based entirely on consideration of the worst that can happen.

In considering our average loss table, we discarded some strategies as being "dominated" by other procedures. Those we rejected are called *inadmissible* strategies. Strategies which are not dominated are called *admissible.*

In our example it might turn out that s_1 or s_2 would be dominated by one of the 76 strategies which have not been examined; on the other hand, other strategies not dominated by s_1 or s_2 might be found. An interesting problem in the theory of decision making is that of finding all the *admissible* strategies.

Certain questions suggest themselves. For example, one may ask why we put so much dependence on the "average losses." This question will be discussed in detail in Chapter 4 on utility. Another question that could be raised would be concerned with the reality of our assumptions. One would actually expect that peak loads of families could vary continuously from less than 15 amp to more than 30 amp. Does our simplification (which was presumably based on previous experience) lead to the adoption of strategies which are liable to have very poor consequences (large losses)? Do you believe that the assumption that the only possible observations are z_1, z_2, z_3, and z_4 is a serious one? Finally, suppose that several observations were available, i.e., the contracter could interview all the members of the family separately. What would be the effect of such data? First, it is clearly apparent that, with the resulting increase in the number of possible combinations of data, the number of strategies available would increase considerably. Second, in statistical problems, the intelligent use of more data generally tends to decrease the average losses.

In this example we ignored the possibility that the strategy could suggest (1) compiling more data before acting, or (2) the use of altogether different data such as examining the number of electric devices in the family's kitchen.

3. PRINCIPLES USED IN DECISION MAKING

Certain points have been illustrated in the example. One is that the main gap in present-day statistical philosphy involves the question of what constitutes a good criterion for selecting strategies. The awareness of this gap permits us to see that in many cases it really is not serious. First, in many industrial applications, the frequencies with which the state of nature is θ_1, θ_2, etc., is approximately known, and one can average the average losses as suggested in the example. In many other applications, the minimum of the maximum average loss is so low that the use of the minimax rule cannot be much improved upon.

Another point is that the statistician must consider the *consequences* of his actions and strategies in order to select a good rule for decision making (strategy). This will be illustrated further in Exercise 1.1 where it is seen that two reasonable people with different loss tables may react differently to the same data even though they apply the same criterion.

Finally, the example illustrates the relation between statistical theory and the scientific method. Essentially every scientist who designs experiments and presents conclusions is engaging in decision making where the costs involved are the time and money for his experiments, on one hand, and damage to society and his reputation if his conclusions are seriously in error, on the other hand. It is not uncommon to hear of nonsense about conclusions being "scientifically proved." In real life very little can be certain. If we tossed a coin a million times, we would not know the exact probability of its falling heads, and we could (although it is unlikely) have a large error in our estimate of this probability. It is true that, generally, scientists attach a large loss to the act of presenting a conclusion which is in error and, hence, they tend to use strategies which decide to present conclusions only when the evidence in favor is very great.

4. SUMMARY

The essential components in decision-making problems are the following :

1. *The available actions* α_1, α_2, $\ldots$. A problem exists when there is a choice of alternative actions. The consequence of taking one of these actions must depend on the state of nature. Usually the difficulty in deciding which action to take is due to the fact that it is not known which of

2. *the possible states of nature* θ_1, θ_2, $\ldots$ is the true one.

3. *The loss table* (*consequence of actions*) measures the cost of taking actions α_1, α_2, $\ldots$ respectively when the states of nature are θ_1, θ_2, $\ldots$ respectively.

Given the loss table, it would be easy to select the best action if the state of nature were known. In other words, the *state of nature* represents the underlying " facts of life," knowledge of which we would like in order to choose a proper action. A