

02125
B498
E2

8964166

James O. Berger

Statistical Decision Theory and Bayesian Analysis

Second Edition



E8964166

With 23 Illustrations



Springer-Verlag
World Publishing Corporation

James O. Berger
Department of Statistics
Purdue University
West Lafayette, IN 47907
U.S.A.

AMS Classification: 60CXX

Reprinted by World Publishing Corporation, Beijing, 1988
for distribution and sale in The People's Republic of China only

Library of Congress Cataloging in Publication Data
Berger, James O.

Statistical decision theory and Bayesian analysis.
(Springer series in statistics)

Bibliography: p.

Includes index.

1. Statistical decision. 2. Bayesian
statistical decision theory. I. Title. II. Series.
QA279.4.B46 1985 519.5'42 85-9891

This is the second edition of *Statistical Decision Theory: Foundations, Concepts, and Methods*. © 1980 Springer-Verlag New York Inc.

© 1980, 1985 by Springer-Verlag New York Inc.

All rights reserved. No part of this book may be translated or reproduced in any form without written permission from Springer-Verlag, 175 Fifth Avenue, New York, New York 10010, U.S.A.

Typeset by J. W. Arrowsmith Ltd., Bristol, England.

Printed and bound by R. R. Donnelley & Sons, Harrisonburg, Virginia.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 7-5062-0262-X

ISBN 0-387-96098-8 Springer-Verlag New York Berlin Heidelberg Tokyo

ISBN 3-540-96098-8 Springer-Verlag Berlin Heidelberg New York Tokyo

Springer Series in Statistics

Advisors:

D. Brillinger, S. Fienberg, J. Gani, J. Hartigan

J. Kiefer, K. Krickeberg

- R. J. Anderson, *Computing in Statistical Science*, xviii, 452 pages, 1984.
- J. O. Berger, *Statistical Decision Making: Principles and Methods*, xviii, 442 pages, 1984.
- R. Brainerd, *Point Processes and Queues: Markovian Systems*, xii, 304 pages, 1981.
- K. Dababneh, *Parameter Estimation and Hypothesis Testing in General Analysis of Stationary Time Series*, xv, 300 pages, 1984.
- R. H. Farrell, *Multivariate Calculus*, xvi, 407 pages, 1983.
- J. A. Goodman, W. H. Krieger, *Measures of Association for Cross Classifications*, x, 146 pages, 1979.
- J. A. Hartigan, *Bayes Theory*, xii, 142 pages, 1983.
- H. Heyer, *Theory of Statistical Experiments*, x, 248 pages, 1982.
- M. Kres, *Statistical Tables for Multivariate Analysis*, xvi, 504 pages, 1982.
- M. R. Leadbetter, G. Lindgren and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, xii, 330 pages, 1983.
- R. G. Miller, Jr., *Simultaneous Statistical Inference*, 2nd edition, xvi, 390 pages, 1981.
- F. Mosteller and D. S. Wallace, *Applied Bayesian and Classical Inference: The Case of the Ferguson Papers*, xxxv, 301 pages, 1983.
- D. Pollard, *Convergence of Stochastic Processes*, xiv, 410 pages, 1984.
- J. W. Pratt and J. D. Gibbons, *Concepts of Nonparametric Theory*, xvi, 405 pages, 1981.
- J. Sachs, *Applied Statistics: A Handbook of Techniques*, 2nd edition, xviii, 700 pages, 1984.
- E. Seneta, *Non-negative Matrices and Markov Chains*, xv, 272 pages, 1981.
- O. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*, xii, 272 pages, 1985.
- V. Vapnik, *Estimation of Dependences Based on Empirical Data*, xvi, 308 pages, 1982.
- K. M. Wolfert, *Introduction to Variance Estimation*, xii, 428 pages, 1985.

Springer Series in Statistics

- D. F. Andrews and A. M. Herzberg, *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. xx, 442 pages, 1985.
- F. J. Anscombe, *Computing in Statistical Science through APL*. xvi, 426 pages, 1981.
- J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. xiv, 425 pages, 1985.
- P. Brémaud, *Point Processes and Queues: Martingale Dynamics*. xviii, 354 pages, 1981.
- K. Dzharidze, *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. xii, 300 pages, 1985.
- R. H. Farrell, *Multivariate Calculation*. xvi, 367 pages, 1985.
- L. A. Goodman and W. H. Kruskal, *Measures of Association for Cross Classifications*. x, 146 pages, 1979.
- J. A. Hartigan, *Bayes Theory*. xii, 145 pages, 1983.
- H. Heyer, *Theory of Statistical Experiments*. x, 289 pages, 1982.
- M. Kres, *Statistical Tables for Multivariate Analysis*. xxii, 504 pages, 1983.
- M. R. Leadbetter, G. Lindgren and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*. xii, 336 pages, 1983.
- R. G. Miller, Jr., *Simultaneous Statistical Inference*, 2nd edition. xvi, 299 pages, 1981.
- F. Mosteller and D. S. Wallace, *Applied Bayesian and Classical Inference: The Case of The Federalist Papers*. xxxv, 301 pages, 1984.
- D. Pollard, *Convergence of Stochastic Processes*. xiv, 215 pages, 1984.
- J. W. Pratt and J. D. Gibbons, *Concepts of Nonparametric Theory*. xvi, 462 pages, 1981.
- L. Sachs, *Applied Statistics: A Handbook of Techniques*, 2nd edition. xxviii, 706 pages, 1984.
- E. Seneta, *Non-Negative Matrices and Markov Chains*. xv, 279 pages, 1981.
- D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. xii, 272 pages, 1985.
- V. Vapnik, *Estimation of Dependences Based on Empirical Data*. xvi, 399 pages, 1982.
- K. M. Wolter, *Introduction to Variance Estimation*. xii, 428 pages, 1985.

Preface

Statistical decision theory and Bayesian analysis are related at a number of levels. First, they are both needed to solve real decision problems, each embodying a description of one of the key elements of a decision problem. At a deeper level, Bayesian analysis and decision theory provide unified outlooks towards statistics; they give a foundational framework for thinking about statistics and for evaluating proposed statistical methods.

The relationships (both conceptual and mathematical) between Bayesian analysis and statistical decision theory are so strong that it is somewhat unnatural to learn one without the other. Nevertheless, major portions of each have developed separately. On the Bayesian side, there is an extensively developed Bayesian theory of statistical inference (both subjective and objective versions). This theory recognizes the importance of viewing statistical analysis conditionally (i.e., treating observed data as known rather than unknown), even when no loss function is to be incorporated into the analysis. There is also a well-developed (frequentist) decision theory, which avoids formal utilization of prior distributions and seeks to provide a foundation for frequentist statistical theory. Although the central thread of the book will be Bayesian decision theory, both Bayesian inference and non-Bayesian decision theory will be extensively discussed. Indeed, the book is written so as to allow, say, the teaching of a course on either subject separately.

Bayesian analysis and, especially, decision theory also have split personalities with regard to their practical orientation. Both can be discussed at a very practical level, and yet they also contain some of the most difficult and elegant theoretical developments in statistics. The book contains a fair amount of material of each type. There is extensive discussion on how to actually do Bayesian decision theory and Bayesian inference, including how

to construct prior distributions and loss functions, as well as how to utilize them. At the other extreme, introductions are given to some of the beautiful theoretical developments in these areas.

The statistical level of the book is formally rather low, in that previous knowledge of Bayesian analysis, decision theory, or advanced statistics is unnecessary. The book will probably be rough going, however, for those without previous exposure to a moderately serious statistics course. For instance, previous exposure to such concepts as sufficiency is desirable. It should also be mentioned that parts of the book are philosophically very challenging; the extreme disagreements that exist among statisticians, concerning the correct approach to statistics, suggest that these fundamental issues are conceptually difficult. Periodic rereading of such material (e.g., Sections 1.6, 4.1, and 4.12), as one proceeds through the book, is recommended.

The mathematical level of the book is, for the most part, at an easy advanced calculus level. Some knowledge of probability is required; at least, say, a knowledge of expectations and conditional probability. From time to time (especially in later chapters) some higher mathematical facts will be employed, but knowledge of advanced mathematics is not required to follow most of the text. Because of the imposed mathematical limitations, some of the stated theorems need, say, additional measurability conditions to be completely precise. Also, less important (but nonignorable) technical conditions for some developments are sometimes omitted, but such developments are called "Results," rather than "Theorems."

The book is primarily concerned with discussing basic issues and principles of Bayesian analysis and decision theory. No systematic attempt is made to present a survey of actual developed methodology, i.e., to present specific developments of these ideas in particular areas of statistics. The examples that are given tend to be rather haphazard, and, unfortunately, do not cover some of the more difficult areas of statistics, such as nonparametrics. Nevertheless, a fair amount of methodology ends up being introduced, one way or another.

This second edition of the book has undergone a title change, with the addition of "Bayesian Analysis." This reflects the major change in the book, namely an extensive upgrading of the Bayesian material, to the point where the book can serve as a text on Bayesian analysis alone. The motivation for this upgrading was the realization that, although I professed to be a "rabid Bayesian" in the first edition (and still am), the first edition was not well suited for a primarily Bayesian course; in particular, it did not highlight the conditional Bayesian perspective properly. In attempting to correct this problem, I fell into the usual revision trap of being unable to resist adding substantial new material on subjects crucial to Bayesian analysis, such as hierarchical Bayes theory, Bayesian calculation, Bayesian communication, and combination of evidence.

For those familiar with the old book, the greatest changes are in Chapters 3 and 4, which were substantially enlarged and almost completely rewritten. Some sections of Chapter 1 were redone (particularly 1.6), and some small subsections were added to Chapter 2. The only significant change to Chapter 5 was the inclusion of an introduction to the now vast field of minimax multivariate estimation (Stein estimation); this has become by far the largest statistical area of development within minimax theory. Only very minor changes were made to Chapter 6, and Chapter 7 was changed only by the addition of a section discussing the issue of optional stopping. A number of changes were made to Chapter 8, in light of recent developments, but no thorough survey was attempted.

In general, no attempt was made to update references in parts of the book that were not rewritten. This, unfortunately, perpetuated a problem with the first edition, namely the lack of references to the early period of decision theory. Many of the decision-theoretic ideas and concepts seem to have become part of the folklore, and I apologize for not making the effort to trace them back to their origins and provide references.

In terms of teaching, the book can be used as a text for a variety of courses. The easiest such use is as a text in a two-semester or three-quarter course on Bayesian analysis and statistical decision theory; one can simply proceed through the book. (Chapters 1 through 4 should take the first semester, and Chapters 5 through 8 the second.) The following are outlines for various possible *single-semester* courses. The first outline is for a master's level course, and has a more applied orientation, while the other outlines also include theoretical material perhaps best suited for Ph.D. students. Of course, quite different arrangements could also be used successfully.

Bayesian Analysis and Decision Theory (Applied)

1 (except 1.4, 1.7, 1.8); 2; 3 (except 3.4, 3.5.5, 3.5.6, 3.5.7); 4 (except 4.4.4, 4.7.4 through 4.7.11, 4.8, 4.11); 7 (except 7.4.2 through 7.4.10, 7.5, 7.6); valuable other material to cover, if there is time, includes 4.7.4, 4.7.5, 4.7.9, 4.7.10, 4.7.11, and 4.11.

Bayesian Analysis and Decision Theory (More Theoretical)

1; 2 (except 2.3, 2.4.3, 2.4.4, 2.4.5); 3 (except 3.4, 3.5.5, 3.5.6, 3.5.7); 4 (except 4.4.4, 4.5.3, 4.6.3, 4.6.4, 4.7.4, 4.7.6, 4.7.7, 4.7.9, 4.7.10, 4.8.3, 4.9, 4.10, 4.11);

- (i) With Minimax Option: 5 (except 5.2.3); parts of 8.
- (ii) With Invariance Option: 6; parts of 8.
- (iii) With Sequential Option: 7 (except 7.4.7 through 7.4.10, 7.5.5, 7.6); parts of 8.

A Mainly Bayesian Course (More Theoretical)

1 (except 1.4, 1.8); 2 (except 2.3); 3 (except 3.5.5 and 3.5.6); 4 (except 4.7.6, 4.7.7); 7 (except 7.4.2 through 7.4.10, 7.5, 7.6); more sequential Bayes could be covered if some of the earlier sections were eliminated.

A Mainly Decision Theory Course (Very Theoretical)

1 (except 1.6); 2 (except 2.3); Sections 3.3, 4.1, 4.2, 4.4, 4.8; 5 (except 5.2.3); 6; 7 (except 7.2, 7.4, 7.7); 8.

I am very grateful to a number of people who contributed, in one way or another, to the book. Useful comments and discussion were received from many sources; particularly helpful were Eric Balder, Mark Berliner, Don Berry, Sudip Bose, Lawrence Brown, Arthur Cohen, Persi Diaconis, Roger Farrell, Leon Gleser, Bruce Hill, Tzou Wu-Jien Joe, T. C. Kao, Jack Kiefer, Sudhakar Kunte, Erich Lehmann, Carl Morris, Herman Rubin, S. Sivaganesan, Bill Studden, Don Wallace, Robert Wolpert, and Arnold Zellner. I am especially grateful to Herman Rubin; he provided most of the material in Subsections 7.4.8 and 7.4.9, and was my "foolishness filter" on much of the rest of the book.

The first edition of the book was typed by Lou Anne Scott, Norma Lucas, Kathy Woods, and Carolyn Knutsen, to all of whom I am very grateful. The highly trying job of typing this revision was undertaken by Norma Lucas, and her skill and cheer throughout the process were deeply appreciated. Finally, I would like to express my appreciation to the John Simon Guggenheim Memorial Foundation, the Alfred P. Sloan Foundation, and the National Science Foundation for support during the writing of the book.

West Lafayette, Indiana
March 1985

JAMES BERGER

Contents



CHAPTER 1

Basic Concepts

1.1	Introduction	1
1.2	Basic Elements	3
1.3	Expected Loss, Decision Rules, and Risk	8
1.3.1	Bayesian Expected Loss	8
1.3.2	Frequentist Risk	9
1.4	Randomized Decision Rules	12
1.5	Decision Principles	16
1.5.1	The Conditional Bayes Decision Principle	16
1.5.2	Frequentist Decision Principles	16
1.6	Foundations	20
1.6.1	Misuse of Classical Inference Procedures	20
1.6.2	The Frequentist Perspective	22
1.6.3	The Conditional Perspective	24
1.6.4	The Likelihood Principle	27
1.6.5	Choosing a Paradigm or Decision Principle	33
1.7	Sufficient Statistics	35
1.8	Convexity	38
	Exercises	41

CHAPTER 2

Utility and Loss

2.1	Introduction	46
2.2	Utility Theory	47
2.3	The Utility of Money	53
2.4	The Loss Function	57
2.4.1	Development from Utility Theory	57

2.4.2	Certain Standard Loss Functions	60
2.4.3	For Inference Problems	64
2.4.4	For Predictive Problems	66
2.4.5	Vector Valued Loss Functions	68
2.5	Criticisms	69
	Exercises	70
 CHAPTER 3		
	Prior Information and Subjective Probability	74
3.1	Subjective Probability	74
3.2	Subjective Determination of the Prior Density	77
3.3	Noninformative Priors	82
3.3.1	Introduction	82
3.3.2	Noninformative Priors for Location and Scale Problems	83
3.3.3	Noninformative Priors in General Settings	87
3.3.4	Discussion	89
3.4	Maximum Entropy Priors	90
3.5	Using the Marginal Distribution to Determine the Prior	94
3.5.1	The Marginal Distribution	94
3.5.2	Information About m	95
3.5.3	Restricted Classes of Priors	97
3.5.4	The ML-II Approach to Prior Selection	99
3.5.5	The Moment Approach to Prior Selection	101
3.5.6	The Distance Approach to Prior Selection	103
3.5.7	Marginal Exchangeability	104
3.6	Hierarchical Priors	106
3.7	Criticisms	109
3.8	The Statistician's Role	113
	Exercises	113
 CHAPTER 4		
	Bayesian Analysis	118
4.1	Introduction	118
4.2	The Posterior Distribution	126
4.2.1	Definition and Determination	126
4.2.2	Conjugate Families	130
4.2.3	Improper Priors	132
4.3	Bayesian Inference	132
4.3.1	Estimation	133
4.3.2	Credible Sets	140
4.3.3	Hypothesis Testing	145
4.3.4	Predictive Inference	157
4.4	Bayesian Decision Theory	158
4.4.1	Posterior Decision Analysis	158
4.4.2	Estimation	161
4.4.3	Finite Action Problems and Hypothesis Testing	163
4.4.4	With Inference Losses	166

4.5	Empirical Bayes Analysis	167
4.5.1	Introduction	167
4.5.2	PEB For Normal Means—The Exchangeable Case	169
4.5.3	PEB For Normal Means—The General Case	173
4.5.4	Nonparametric Empirical Bayes Analysis	178
4.6	Hierarchical Bayes Analysis	180
4.6.1	Introduction	180
4.6.2	For Normal Means—The Exchangeable Case	183
4.6.3	For Normal Means—The General Case	190
4.6.4	Comparison with Empirical Bayes Analysis	193
4.7	Bayesian Robustness	195
4.7.1	Introduction	195
4.7.2	The Role of the Marginal Distribution	199
4.7.3	Posterior Robustness: Basic Concepts	203
4.7.4	Posterior Robustness: ε -Contamination Class	206
4.7.5	Bayes Risk Robustness and Use of Frequentist Measures	213
4.7.6	Gamma-Minimax Approach	215
4.7.7	Uses of the Risk Function	218
4.7.8	Some Robust and Nonrobust Situations	223
4.7.9	Robust Priors	228
4.7.10	Robust Priors for Normal Means	236
4.7.11	Other Issues in Robustness	247
4.8	Admissibility of Bayes Rules and Long Run Evaluations	253
4.8.1	Admissibility of Bayes Rules	253
4.8.2	Admissibility of Generalized Bayes Rules	254
4.8.3	Inadmissibility and Long Run Evaluations	257
4.9	Bayesian Calculation	262
4.9.1	Numerical Integration	262
4.9.2	Monte Carlo Integration	263
4.9.3	Analytic Approximations	265
4.10	Bayesian Communication	267
4.10.1	Introduction	267
4.10.2	An Illustration: Testing a Point Null Hypothesis	268
4.11	Combining Evidence and Group Decisions	271
4.11.1	Combining Probabilistic Evidence	272
4.11.2	Combining Decision-Theoretic Evidence	277
4.11.3	Group Decision Making	278
4.12	Criticisms	281
4.12.1	Non-Bayesian Criticisms	281
4.12.2	Foundational Criticisms	283
	Exercises	286

CHAPTER 5

Minimax Analysis

5.1	Introduction	308
5.2	Game Theory	310
5.2.1	Basic Elements	310
5.2.2	General Techniques for Solving Games	319

5.2.3	Finite Games	325
5.2.4	Games with Finite Θ	331
5.2.5	The Supporting and Separating Hyperplane Theorems	339
5.2.6	The Minimax Theorem	345
5.3	Statistical Games	347
5.3.1	Introduction	347
5.3.2	General Techniques for Solving Statistical Games	349
5.3.3	Statistical Games with Finite Θ	354
5.4	Classes of Minimax Estimators	359
5.4.1	Introduction	359
5.4.2	The Unbiased Estimator of Risk	361
5.4.3	Minimax Estimators of a Normal Mean Vector	363
5.4.4	Minimax Estimators of Poisson Means	369
5.5	Evaluation of the Minimax Principle	370
5.5.1	Admissibility of Minimax Rules	371
5.5.2	Rationality and the Minimax Principle	371
5.5.3	Comparison with the Bayesian Approach	373
5.5.4	The Desire to Act Conservatively	376
5.5.5	Minimax Regret	376
5.5.6	Conclusions	378
	Exercises	379

CHAPTER 6

Invariance	388
6.1 Introduction	388
6.2 Formulation	391
6.2.1 Groups of Transformations	391
6.2.2 Invariant Decision Problems	393
6.2.3 Invariant Decision Rules	395
6.3 Location Parameter Problems	397
6.4 Other Examples of Invariance	400
6.5 Maximal Invariants	402
6.6 Invariance and Noninformative Priors	406
6.6.1 Right and Left Invariant Haar Densities	406
6.6.2 The Best Invariant Rule	409
6.6.3 Confidence and Credible Sets	414
6.7 Invariance and Minimality	418
6.8 Admissibility of Invariant Rules	422
6.9 Conclusions	423
Exercises	425

CHAPTER 7

Preposterior and Sequential Analysis	432
7.1 Introduction	432
7.2 Optimal Fixed Sample Size	435
7.3 Sequential Analysis—Notation	441

7.4	Bayesian Sequential Analysis	442
7.4.1	Introduction	442
7.4.2	Notation	445
7.4.3	The Bayes Decision Rule	446
7.4.4	Constant Posterior Bayes Risk	447
7.4.5	The Bayes Truncated Procedure	448*
7.4.6	Look Ahead Procedures	455
7.4.7	Inner Truncation	459
7.4.8	Approximating the Bayes Procedure and the Bayes Risk	462
7.4.9	Theoretical Results	467
7.4.10	Other Techniques for Finding a Bayes Procedure	473
7.5	The Sequential Probability Ratio Test	481
7.5.1	The SPRT as a Bayes Procedure	482
7.5.2	Approximating the Power Function and the Expected Sample Size	485
7.5.3	Accuracy of the Wald Approximations	495
7.5.4	Bayes Risk and Admissibility	498
7.5.5	Other Uses of the SPRT	500
7.6	Minimax Sequential Procedures	501
7.7	The Evidential Relevance of the Stopping Rule	502
7.7.1	Introduction	502
7.7.2	The Stopping Rule Principle	502
7.7.3	Practical Implications	504
7.7.4	Criticisms of the Stopping Rule Principle	506
7.7.5	Informative Stopping Rules	510
7.8	Discussion of Sequential Loss Functions	511
	Exercises	513
CHAPTER 8		
	Complete and Essentially Complete Classes	521
8.1	Preliminaries	521
8.2	Complete and Essentially Complete Classes from Earlier Chapters	522
8.2.1	Decision Rules Based on a Sufficient Statistic	522
8.2.2	Nonrandomized Decision Rules	523
8.2.3	Finite Θ	523
8.2.4	The Neyman-Pearson Lemma	523
8.3	One-Sided Testing	525
8.4	Monotone Decision Problems	530
8.4.1	Monotone Multiple Decision Problems	530
8.4.2	Monotone Estimation Problems	534
8.5	Limits of Bayes Rules	537
8.6	Other Complete and Essentially Complete Classes of Tests	538
8.6.1	Two-Sided Testing	538
8.6.2	Higher Dimensional Results	538
8.6.3	Sequential Testing	540
8.7	Complete and Essentially Complete Classes in Estimation	541
8.7.1	Generalized Bayes Estimators	541
8.7.2	Identifying Generalized Bayes Estimators	543

8.8	Continuous Risk Functions	544
8.9	Proving Admissibility and Inadmissibility	546
8.9.1	Stein's Necessary and Sufficient Condition for Admissibility	546
8.9.2	Proving Admissibility	547
8.9.3	Proving Inadmissibility	550
8.9.4	Minimal (or Nearly Minimal) Complete Classes	552
	Exercises	554
APPENDIX 1		
	Common Statistical Densities	559
I	Continuous	559
II	Discrete	562
APPENDIX 2		
	Supplement to Chapter 4	563
I	Definition and Properties of H_m	563
II	Development of (4.121) and (4.122)	564
III	Verification of Formula (4.123)	565
APPENDIX 3		
	Technical Arguments from Chapter 7	568
I	Verification of Formula (7.8)	568
II	Verification of Formula (7.10)	569
	Bibliography	571
	Notation and Abbreviations	599
	Author Index	603
	Subject Index	609

CHAPTER 1

Basic Concepts

1.1. Introduction

Decision theory, as the name implies, is concerned with the problem of making decisions. Statistical decision theory is concerned with the making of decisions in the presence of statistical knowledge which sheds light on some of the uncertainties involved in the decision problem. We will, for the most part, assume that these uncertainties can be considered to be unknown numerical quantities, and will represent them by θ (possibly a vector or matrix).

As an example, consider the situation of a drug company deciding whether or not to market a new pain reliever. Two of the many factors affecting its decision are the proportion of people for which the drug will prove effective (θ_1), and the proportion of the market the drug will capture (θ_2). Both θ_1 and θ_2 will be generally unknown, though typically experiments can be conducted to obtain statistical information about them. This problem is one of decision theory in that the ultimate purpose is to decide whether or not to market the drug, how much to market, what price to charge, etc.

Classical statistics is directed towards the use of sample information (the data arising from the statistical investigation) in making inferences about θ . These classical inferences are, for the most part, made without regard to the use to which they are to be put. In decision theory, on the other hand, an attempt is made to combine the sample information with other relevant aspects of the problem in order to make the best decision.

In addition to the sample information, two other types of information are typically relevant. The first is a knowledge of the possible consequences of the decisions. Often this knowledge can be quantified by determining the loss that would be incurred for each possible decision and for the various

possible values of θ . (Statisticians seem to be pessimistic creatures who think in terms of losses. Decision theorists in economics and business talk instead in terms of gains (utility). As our orientation will be mainly statistical, we will use the loss function terminology. Note that a gain is just a negative loss, so there is no real difference between the two approaches.)

The incorporation of a loss function into statistical analysis was first studied extensively by Abraham Wald; see Wald (1950), which also reviews earlier work in decision theory.

In the drug example, the losses involved in deciding whether or not to market the drug will be complicated functions of θ_1 , θ_2 , and many other factors. A somewhat simpler situation to consider is that of estimating θ_1 , for use, say, in an advertising campaign. The loss in underestimating θ_1 arises from making the product appear worse than it really is (adversely affecting sales), while the loss in overestimating θ_1 would be based on the risks of possible penalties for misleading advertising.

The second source of nonsample information that is useful to consider is called prior information. This is information about θ arising from sources other than the statistical investigation. Generally, prior information comes from past experience about similar situations involving similar θ . In the drug example, for instance, there is probably a great deal of information available about θ_1 and θ_2 from different but similar pain relievers.

A compelling example of the possible importance of prior information was given by L. J. Savage (1961). He considered the following three statistical experiments:

1. A lady, who adds milk to her tea, claims to be able to tell whether the tea or the milk was poured into the cup first. In all of ten trials conducted to test this, she correctly determines which was poured first.
2. A music expert claims to be able to distinguish a page of Haydn score from a page of Mozart score. In ten trials conducted to test this, he makes a correct determination each time.
3. A drunken friend says he can predict the outcome of a flip of a fair coin. In ten trials conducted to test this, he is correct each time.

In all three situations, the unknown quantity θ is the probability of the person answering correctly. A classical significance test of the various claims would consider the null hypothesis (H_0) that $\theta = 0.5$ (i.e., the person is guessing). In all three situations this hypothesis would be rejected with a (one-tailed) significance level of 2^{-10} . Thus the above experiments give strong evidence that the various claims are valid.

In situation 2 we would have no reason to doubt this conclusion. (The outcome is quite plausible with respect to our prior beliefs.) In situation 3, however, our prior opinion that this prediction is impossible (barring a belief in extrasensory perception) would tend to cause us to ignore the experimental evidence as being a lucky streak. In situation 1 it is not quite clear what to think, and different people will draw different conclusions