

# Decision Theory: Principles and approaches

Giovanni PARMIGIANI  
Johns Hopkins University, Baltimore, MD, USA

Lurdes Yoshiko Tani INOUE  
University of Washington, Seattle, WA, USA

with contributions from Hedibert Lopes  
University of Chicago, Chicago, IL, USA

JOHN WILEY & SONS  
Chichester · New York · Brisbane · Toronto · Singapore

to our advisors: Don Berry, Morrie De Groot and Jay Kadane,  
and to their advisors: Jay Kadane, Jimmy Savage and Herman Chernoff.

# Preface

## Goals

The goal of this book is to give an overview of fundamental ideas and results about rational decision making under uncertainty, highlighting the implications of these results for the philosophy and practice of statistics. The book grew from lecture notes from graduate courses taught at the Institute of Statistics and Decision Sciences at Duke, at the Johns Hopkins School of Public Health, and at the University of Washington. It is designed primarily for graduate students in statistics and biostatistics, both at the master and PhD level. However, the interdisciplinary nature of the material should make it interesting to students and researchers in economics (choice theory, econometrics), engineering (signal processing; risk analysis), computer science (pattern recognition and artificial intelligence), and scientists who are interested in the general principles of experimental design and analysis.

Rational decision making has been a chief area of investigation in a number of disciplines, in some cases for centuries. Several of the contributions and viewpoints are relevant to both the education of a well rounded statistician and to the development of sound statistical practices. Because of the wealth of important ideas, and the pressure from competing needs in current statistical curricula, our first course in decision theory aims for breadth rather than depth. We paid special attention to two aspects: bridging the gaps among the different fields that have contributed to rational decision making, and presenting ideas in a unified framework and notation while respecting and highlighting the different and sometimes conflicting perspectives.

With this in mind, we felt that a standard textbook format would be too constraining for us and not sufficiently stimulating for the students. So our approach has been to write a “tour guide” to some of the ideas and papers that have contributed to making decision theory so fascinating and important. We selected a set of exciting papers and book chapters, and developed a self contained lecture around each one. Some lectures are close to the source, while other stray far from their original inspiration. Naturally, many important articles have been left out of the tour. Our goal was to select a set that would work well together in conveying an overall view of the fields and controversies.

We decided to cover three areas: the axiomatic foundations of decision theory; statistical decision theory; and optimal design of experiments. At many universities, these are the subject of separate courses, often taught in different departments and schools. Current curricula in statistics and biostatistics are increasingly emphasizing interdisciplinary training, reflecting similar trends in research. Our plan reflects this need. We also hope to contribute to increased interaction among the disciplines by training students to appreciate the differences and similarities among the approaches.

We designed our tour of decision theoretic ideas so that students might emerge with their own overall philosophy of decision making and statistics. Ideally that philosophy will be the result of contact with some of the key ideas and controversies in the different fields. We attempted to put contributions of each article in some historical perspective and to highlight developments that followed. We also developed a consistent unified notation for the entire material and emphasized the relationships among different disciplines and points of view. Most lectures include current day materials, methods and results, and try at the same time to preserve the viewpoint and flavor of the original contributions.

With few exceptions, the mathematical level of the book is basic. Advanced calculus and intermediate statistical inference are useful prerequisites, but an enterprising student can profit from most of the the book even without this background. The challenging aspect of the book lies in the swift pace at which each lecture introduces new and different concepts and points of view.

Some lectures have grown beyond the size that can be delivered during a one-hour-and-a-half session. Some others merge materials that were often taught as two separate lectures. But for the most part, the lecture-session correspondence should work reasonably well. The style is also closer to that of transcribed lecture notes than that of a treatise. Each lecture is completed by worked examples and exercises that have been helpful to us in teaching this material. Many proofs, easy and hard, are left to the student.

### **Acknowledgments**

We have intellectual debt to more people than we can list, but a special place in this list is occupied by courses we took and lecture notes we read. Giovanni's course at Duke was initially developed from two main sources. The first are lectures from Teddy Seidenfeld's course on the Foundations of Statistics. Giovanni only took it five times—he will firmly hold he did not choose the stopping rule: left to his own devices he would have taken that class forever. The second are lectures from Schervish's course on Advanced Statistics, from which his book on the "Theory of Statistics" would ultimately develop. We also had access to a very insightful bootleg of Charles Stein's lecture notes at Stanford, from an edition of the course taught by Persi Diaconis.

Dennis Lindley reviewed an early draft and gave very constructive

comments and encouragement. Other anonymous reviewers gave helpful feedback. Bruno Sansó used our notes to teach his class at the University of California at Santa Cruz, and gave us detailed comments. We used our notes in teaching for over a decade. Many students braved earlier drafts, gave useful feedback through questions, conversations, solutions to problems, and sometimes highly informative puzzled looks. Marty Macintosh shared his precious correspondence with Herman Chernoff.

Both of us are grateful to Hedibert Lopes, with whom our long journey to writing this book had started back in the mid nineties. His notes from Giovanni's classes were used extensively in early versions of this book, and some figures, problems, and examples still carry his hallmark.

Lurdes is thankful to Sergio Wechsler who opened the door to new ways of thinking about statistics and introducing her to decision theory. She thanks Giovanni for inviting her to this journey, which through bumps and laughter, has been a lifetime experience. She cannot wait for the next one [well, give and take some time off for her recovery from the thrill!]. She wishes to thank the loving support from her brothers Roberto, Carlos and Silvio and from her uncles Masao, Olinda and Tadazumi. Finally, her loving gratitude goes to her parents, Satie and Kasuo, and her grandmother Matta, for the inspiring memories and lessons that guide Lurdes.

Giovanni still has mixed feelings about the day Marco Scarsini handed him a copy of Wald's book on decision functions, with the assignment of reporting about it to an undergraduate discussion group. Later Michele Cifarelli, Guido Consonni, Morrie DeGroot, Jay Kadane, Teddy Seidenfeld, Mark Schervish, Nick Polson, Don Berry, Pietro Muliere, Peter Müller, and David Matchar fueled his intellectual passion for rational decision making. Giovanni's wife Francesca is a statistician who, despite her impact on national policy making at various levels, is a bit bored by the kind of decision theory her husband favors, and perhaps baffled by the scant influence all the talking about rationality has had on his personal behavior. Nevertheless, she has been fully supportive of this never-ending project, in more ways than one can list. Giovanni thinks working with Lurdes has been absolutely fantastic. He has not told her yet, but he is already thinking about notation changes for the second edition....

December 23, 2008

Giovanni Parmigiani, Lurdes Y. T. Inoue



# Contents

<b>Preface</b> . . . . .	1
Goals . . . . .	1
Acknowledgments . . . . .	2
<b>1 Introduction</b> . . . . .	1
1.1 Controversies . . . . .	1
1.2 A Guided Tour of Decision Theory . . . . .	6
<b>I Foundations</b>	<b>11</b>
<b>2 Coherence</b> . . . . .	13
2.1 The “Dutch Book” Theorem . . . . .	14
2.1.1 Betting odds . . . . .	14
2.1.2 Coherence and the axioms of probability . . . . .	17
2.1.3 Coherent Conditional Probabilities . . . . .	20
2.1.4 The implications of Dutch Book theorems . . . . .	22
2.2 Temporal Coherence . . . . .	24
2.3 Scoring Rules and the axioms of probabilities . . . . .	27
2.4 Exercises . . . . .	27
<b>3 Utility</b> . . . . .	33
3.1 Saint Petersburg Paradox . . . . .	34
3.2 Expected Utility Theory and the Theory of Means . . . . .	37
3.2.1 Utility and Means . . . . .	37
3.2.2 Associative Means . . . . .	38
3.2.3 Functional Means . . . . .	39
3.3 The Expected Utility Principle . . . . .	40
3.4 von Neumann–Morgenstern Representation Theorem . . . . .	42
3.4.1 Axioms . . . . .	42
3.4.2 Representation of Preferences via Expected Utility . . . . .	44
3.5 Allais’ Criticism . . . . .	48

3.6	Extensions . . . . .	50
3.7	Exercises . . . . .	50
<b>4</b>	<b>Utility in Action . . . . .</b>	<b>57</b>
4.1	The “standard gamble” . . . . .	58
4.2	Utility of Money . . . . .	59
4.2.1	Certainty Equivalents . . . . .	59
4.2.2	Risk Aversion . . . . .	59
4.2.3	A Measure of Risk Aversion . . . . .	61
4.3	Utility functions for medical decisions . . . . .	65
4.3.1	Length and Quality of Life . . . . .	65
4.3.2	Standard gamble for health states . . . . .	65
4.3.3	The time trade-off methods . . . . .	66
4.3.4	Relation between QALYs and utilities . . . . .	67
4.3.5	Utilities for time in ill health . . . . .	68
4.3.6	Difficulties in assessing utility . . . . .	70
4.4	Exercises . . . . .	71
<b>5</b>	<b>Ramsey and Savage . . . . .</b>	<b>77</b>
5.1	Ramsey’s Theory . . . . .	78
5.2	Savage’s theory . . . . .	82
5.2.1	Notation and overview . . . . .	82
5.2.2	The sure thing principle . . . . .	84
5.2.3	Conditional and a Posteriori Preferences . . . . .	87
5.2.4	Subjective probability . . . . .	87
5.2.5	Utility and expected utility . . . . .	92
5.3	Allais Revisited . . . . .	93
5.4	Ellsberg Paradox . . . . .	94
5.5	Exercises . . . . .	96
<b>6</b>	<b>State Independence . . . . .</b>	<b>99</b>
6.1	Horse Lotteries . . . . .	100
6.2	State-dependent utilities . . . . .	101
6.3	State-independent utilities . . . . .	103
6.4	Anscombe-Aumann Representation Theorem . . . . .	105
6.5	Exercises . . . . .	107
<b>II</b>	<b>Statistical Decision Theory . . . . .</b>	<b>111</b>
<b>7</b>	<b>Decision Functions . . . . .</b>	<b>113</b>
7.1	Basic Concepts . . . . .	114
7.1.1	The Loss Function . . . . .	114
7.1.2	Minimax . . . . .	116



7.1.3	Expected Utility Principle . . . . .	118
7.1.4	Illustrations . . . . .	119
7.2	Data-based Decisions . . . . .	122
7.2.1	Risk . . . . .	122
7.2.2	Optimality Principles . . . . .	124
7.2.3	Rationality principles and the Likelihood Principle . . .	125
7.2.4	Nuisance Parameters . . . . .	127
7.3	The Travel Insurance Example . . . . .	129
7.4	Randomized Decision Rules . . . . .	135
7.5	Classification and Hypothesis Tests . . . . .	136
7.5.1	Hypothesis Testing . . . . .	136
7.5.2	Multiple Hypothesis Testing . . . . .	139
7.5.3	Classification . . . . .	142
7.6	Estimation . . . . .	143
7.6.1	Point Estimation . . . . .	143
7.6.2	Interval Inference . . . . .	146
7.7	Minimax—Bayes Connections . . . . .	147
7.8	Exercises . . . . .	154
<b>8</b>	<b>Admissibility . . . . .</b>	<b>159</b>
8.1	Admissibility and Completeness . . . . .	160
8.2	Admissibility and Minimax . . . . .	162
8.3	Admissibility and Bayes . . . . .	163
8.3.1	Proper Bayes Rules . . . . .	163
8.3.2	Generalized Bayes Rules . . . . .	164
8.4	Complete Classes . . . . .	168
8.4.1	Completeness and Bayes . . . . .	168
8.4.2	Sufficiency and the Rao–Blackwell Inequality . . . . .	170
8.4.3	The Neyman–Pearson Lemma . . . . .	172
8.5	Using the same $\alpha$ level across studies with different sample sizes is inadmissible. . . . .	173
8.6	Exercises . . . . .	175
<b>9</b>	<b>Shrinkage . . . . .</b>	<b>179</b>
9.1	The Stein Effect . . . . .	180
9.2	Geometric and Empirical Bayes Heuristics . . . . .	183
9.2.1	Is $\mathbf{x}$ too big for $\boldsymbol{\theta}$ ? . . . . .	183
9.2.2	Empirical Bayes Shrinkage . . . . .	185
9.3	General Shrinkage Functions . . . . .	187
9.3.1	Unbiased Estimation of the Risk of $\mathbf{x} + \mathbf{g}(\mathbf{x})$ . . . . .	187
9.3.2	Bayes and Minimax Shrinkage . . . . .	189
9.4	Shrinkage with different Likelihood and Losses . . . . .	191
9.5	Exercises . . . . .	192

<b>10 Scoring Rules</b>	195
10.1 Betting and Forecasting	196
10.2 Scoring Rules	197
10.2.1 Definition	197
10.2.2 Proper Scoring Rules	198
10.2.3 The quadratic scoring rules	199
10.2.4 Scoring rules that are not proper	200
10.3 Local Scoring Rules	201
10.4 Calibration and Refinement	204
10.4.1 The well-calibrated forecaster	204
10.4.2 Are Bayesians Well Calibrated?	209
10.5 Exercises	212
<b>11 Choosing Models</b>	215
11.1 The “True Model” Perspective	216
11.1.1 Model Probabilities	216
11.1.2 Model Selection and Bayes Factors	218
11.1.3 Model averaging for prediction and selection	220
11.2 Model Elaborations	222
11.3 Exercises	225
<b>III Optimal Design</b>	<b>227</b>
<b>12 Dynamic Programming</b>	229
12.1 History	230
12.2 The Travel Insurance Example, Revisited	232
12.3 Dynamic Programming	235
12.3.1 Two-stage finite decision problems	235
12.3.2 More than Two Stages	238
12.4 Trading off immediate Gains and Information	240
12.4.1 The Secretary Problem	240
12.4.2 The prophet inequality	243
12.5 Sequential Clinical Trials	245
12.5.1 Two-armed bandit problems	245
12.5.2 Adaptive Designs for Binary Outcomes	246
12.6 Variable selection in multiple regression	248
12.7 Computing	250
12.8 Exercises	253
<b>13 Changes in Utility as Information</b>	263
13.1 Measuring the Value of Information	264
13.1.1 The Value Function	264
13.1.2 Information from a Perfect Experiment	266

13.1.3	Information from a Statistical Experiment . . . . .	267
13.1.4	The Distribution of Information . . . . .	272
13.2	Examples . . . . .	273
13.2.1	Tasting Grapes . . . . .	273
13.2.2	Medical Testing . . . . .	274
13.2.3	Hypothesis Testing . . . . .	282
13.3	Lindley Information . . . . .	285
13.3.1	Definition . . . . .	285
13.3.2	Properties . . . . .	286
13.3.3	Computing . . . . .	288
13.3.4	Optimal Design . . . . .	289
13.4	Minimax and the Value of Information . . . . .	292
13.5	Exercises . . . . .	294
<b>14</b>	<b>Sample Size . . . . .</b>	<b>297</b>
14.1	Decision Theoretic Approaches to Sample Size . . . . .	298
14.1.1	Sample Size and Power . . . . .	298
14.1.2	Sample size as a decision problem . . . . .	298
14.1.3	Bayes and Minimax Optimal Sample Size . . . . .	300
14.1.4	A Minimax Paradox . . . . .	301
14.1.5	Goal Sampling . . . . .	303
14.2	Computing . . . . .	306
14.3	Examples . . . . .	309
14.3.1	Point estimation with quadratic loss . . . . .	309
14.3.2	Composite hypothesis testing . . . . .	310
14.3.3	A Two-Action problem with linear utility . . . . .	312
14.3.4	Lindley Information for Exponential Data . . . . .	315
14.3.5	Multicenter Clinical Trials . . . . .	316
14.4	Exercises . . . . .	318
<b>15</b>	<b>Stopping . . . . .</b>	<b>333</b>
15.1	Historical note . . . . .	334
15.2	A Motivating Example . . . . .	336
15.3	Bayesian Optimal Stopping . . . . .	337
15.3.1	Notation . . . . .	337
15.3.2	Bayes Sequential Procedure . . . . .	339
15.3.3	Bayes Truncated Procedure . . . . .	341
15.4	Examples . . . . .	343
15.4.1	Hypotheses Testing . . . . .	343
15.4.2	An example with equivalence between sequential and fixed sample size designs. . . . .	346
15.5	Sequential sampling to reduce uncertainty . . . . .	347
15.6	The Stopping Rule Principle . . . . .	349
15.6.1	Stopping Rules and the Likelihood Principle . . . . .	349

15.6.2 Sampling to a Foregone Conclusions . . . . .	351
15.7 Exercises . . . . .	353
<b>16 Appendix . . . . .</b>	<b>357</b>
16.1 Notation . . . . .	357
16.2 Relations . . . . .	362
16.3 Probability (density) functions of some distributions . . . . .	364
16.4 Conjugate updating . . . . .	364
REFERENCES . . . . .	366