

---

---

# PREFACE TO THE 2015 EDITION

In this preface, we start with an overview of developments in statistics since the first (1977) edition, then give separate overviews of Volumes I and II of the second edition.

In the last 40 some years statistics has changed enormously under the impact of several forces:

- (1) The generation of what were once unusual types of data such as images, trees (phylogenetic and other), and other types of combinatorial objects.
- (2) The generation of enormous amounts of data—terabytes (the equivalent of  $10^{12}$  characters) for an astronomical survey over three years.
- (3) The possibility of implementing computations of a magnitude that would have once been unthinkable.

The underlying sources of these changes have been the exponential change in computing speed (Moore's "law") and the development of devices (computer controlled) using novel instruments and scientific techniques (e.g., NMR tomography, gene sequencing). These techniques often have a strong intrinsic computational component. Tomographic data are the result of mathematically based processing. Sequencing is done by applying computational algorithms to raw gel electrophoresis data.

As a consequence the emphasis of statistical theory has shifted away from small sample optimality results in a number of directions:

- (1) Methods for inference based on larger numbers of observations and minimal assumptions—asymptotic methods in non- and semiparametric models, models with "infinite" number of parameters.
- (2) The construction of models for time series, temporal spatial series, and other complex data structures using sophisticated probability modeling but again relying for analytical results on asymptotic approximation. Multiparameter models are the rule.
- (3) The use of methods of inference involving simulation as a key element such as the bootstrap and Markov Chain Monte Carlo.

- (4) The development of techniques not describable in “closed mathematical form” but rather through elaborate algorithms for which problems of existence of solutions are important and far from obvious.
- (5) The study of the interplay between numerical and statistical considerations. Despite advances in computing speed, some methods run quickly in real time. Others do not and some though theoretically attractive cannot be implemented in a human lifetime.
- (6) The study of the interplay between the number of observations and the number of parameters of a model and the beginnings of appropriate asymptotic theories.

There have been other important consequences such as the extensive development of graphical and other exploratory methods for which theoretical development and connection with mathematics have been minimal. These will not be dealt with in our work.

In this edition we pursue our philosophy of describing the basic concepts of mathematical statistics relating theory to practice.

## Volume I

This volume presents the basic classical statistical concepts at the Ph.D. level without requiring measure theory. It gives careful proofs of the major results and indicates how the theory sheds light on the properties of practical methods. The topics include estimation, prediction, testing, confidence sets, Bayesian analysis and the more general approach of decision theory.

We include from the start in Chapter 1 non- and semiparametric models, then go to parameters and parametric models stressing the role of identifiability. From the beginning we stress function-valued parameters, such as the density, and function-valued statistics, such as the empirical distribution function. We also, from the start, include examples that are important in applications, such as regression experiments. There is extensive material on Bayesian models and analysis and extended discussion of prediction and  $k$ -parameter exponential families. These objects that are the building blocks of most modern models require concepts involving moments of random vectors and convexity that are given in Appendix B.

Chapter 2 deals with estimation and includes a detailed treatment of maximum likelihood estimates (MLEs), including a complete study of MLEs in canonical  $k$ -parameter exponential families. Other novel features of this chapter include a detailed analysis, including proofs of convergence, of a standard but slow algorithm (coordinate descent) for convex optimization, applied, in particular to computing MLEs in multiparameter exponential families. We also give an introduction to the EM algorithm, one of the main ingredients of most modern algorithms for inference. Chapters 3 and 4 are on the theory of testing and confidence regions, including some optimality theory for estimation as well and elementary robustness considerations.

Chapter 5 is devoted to basic asymptotic approximations with one dimensional parameter models as examples. It includes proofs of consistency and asymptotic normality and optimality of maximum likelihood procedures in inference and a section relating Bayesian and frequentist inference via the Bernstein–von Mises theorem.

Finally, Chapter 6 is devoted to inference in multivariate (multiparameter) models. Included are asymptotic normality and optimality of maximum likelihood estimates, inference in the general linear model, Wilks theorem on the asymptotic distribution of the likelihood ratio test, the Wald and Rao statistics and associated confidence regions, and some parallels to the optimality theory and comparisons of Bayes and frequentist procedures given in the one dimensional parameter case in Chapter 5. Chapter 6 also develops the asymptotic joint normality of estimates that are solutions to estimating equations and presents Huber's Sandwich formula for the asymptotic covariance matrix of such estimates. Generalized linear models, including binary logistic regression, are introduced as examples. Robustness from an asymptotic theory point of view appears also. This chapter uses multivariate calculus in an intrinsic way and can be viewed as an essential prerequisite for the more advanced topics of Volume II.

Volume I includes Appendix A on basic probability and a larger Appendix B, which includes more advanced topics from probability theory such as the multivariate Gaussian distribution, weak convergence in Euclidean spaces, and probability inequalities as well as more advanced topics in matrix theory and analysis. The latter include the principal axis and spectral theorems for Euclidean space and the elementary theory of convex functions on  $R^d$  as well as an elementary introduction to Hilbert space theory. As in the first edition, we do not require measure theory but assume from the start that our models are what we call "regular." That is, we assume either a discrete probability whose support does not depend on the parameter set, or the absolutely continuous case with a density. Hilbert space theory is not needed, but for those who know this topic Appendix B points out interesting connections to prediction and linear regression analysis.

Appendix B is as self-contained as possible with proofs of most statements, problems, and references to the literature for proofs of the deepest results such as the spectral theorem. The reason for these additions are the changes in subject matter necessitated by the current areas of importance in the field.

For the first volume of the second edition we would like to add thanks to Jianging Fan, Michael Jordan, Jianhua Huang, Ying Qing Chen, and Carl Spruill and the many students who were guinea pigs in the basic theory course at Berkeley. We also thank Faye Yeager for typing, Michael Ostland and Simon Cawley for producing the graphs, Yoram Gat for proofreading that found not only typos but serious errors, and Prentice Hall for generous production support.

## Volume II

Volume II of the second edition will be forthcoming in 2015. It presents what we think are some of the most important statistical concepts, methods, and tools developed since the first edition. Topics to be included are: asymptotic efficiency in semiparametric models, semiparametric maximum likelihood estimation, survival analysis including Cox regression, classification, methods of inference based on sieve models, model selection, Monte Carlo methods such as the bootstrap and Markov Chain Monte Carlo, nonparametric curve estimation, and machine learning including support vector machines and classification and regression trees (CART).

The basic asymptotic tools that will be developed or presented, in part in the text and, in part in appendices, are weak convergence for random processes, elementary empirical process theory, and the functional delta method.

With the tools and concepts developed in this second volume students will be ready for advanced research in modern statistics.

We thank Akichika Ozeki and Sören Künzel for pointing out errors and John Kimmel and CRC Press for production support. We also thank Dee Frana and especially Anne Chong who typed 90% of Volume II for word processing.

Last and most important we would like to thank our wives, Nancy Kramer Bickel and Joan H. Fujimura, and our families for support, encouragement, and active participation in an enterprise that at times seemed endless, appeared gratifyingly ended in 1976 but has, with the field, taken on a new life.

Peter J. Bickel  
bickel@stat.berkeley.edu  
Kjell Doksum  
doksum@stat.wisc.edu

---

---

# CONTENTS

## PREFACE TO THE 2015 EDITION

xxi

## 1 STATISTICAL MODELS, GOALS, AND PERFORMANCE CRITERIA

1

### 1.1 Data, Models, Parameters, and Statistics

1

#### 1.1.1 Data and Models

1

#### 1.1.2 Parametrizations and Parameters

6

#### 1.1.3 Statistics as Functions on the Sample Space

8

#### 1.1.4 Examples, Regression Models

9

### 1.2 Bayesian Models

12

### 1.3 The Decision Theoretic Framework

16

#### 1.3.1 Components of the Decision Theory Framework

17

#### 1.3.2 Comparison of Decision Procedures

24

#### 1.3.3 Bayes and Minimax Criteria

26

### 1.4 Prediction

32

### 1.5 Sufficiency

41

### 1.6 Exponential Families

49

#### 1.6.1 The One-Parameter Case

49

#### 1.6.2 The Multiparameter Case

53

#### 1.6.3 Building Exponential Families

56

#### 1.6.4 Properties of Exponential Families

58

#### 1.6.5 Conjugate Families of Prior Distributions

62

### 1.7 Problems and Complements

66

### 1.8 Notes

95

### 1.9 References

96

ix

<b>2</b>	<b>METHODS OF ESTIMATION</b>	<b>99</b>
2.1	Basic Heuristics of Estimation	99
2.1.1	Minimum Contrast Estimates; Estimating Equations	99
2.1.2	The Plug-In and Extension Principles	102
2.2	Minimum Contrast Estimates and Estimating Equations	107
2.2.1	Least Squares and Weighted Least Squares	107
2.2.2	Maximum Likelihood	114
2.3	Maximum Likelihood in Multiparameter Exponential Families	121
2.4	Algorithmic Issues	127
2.4.1	The Method of Bisection	127
2.4.2	Coordinate Ascent	129
2.4.3	The Newton–Raphson Algorithm	131
2.4.4	The EM (Expectation/Maximization) Algorithm	133
2.5	Problems and Complements	138
2.6	Notes	158
2.7	References	159
<b>3</b>	<b>MEASURES OF PERFORMANCE</b>	<b>161</b>
3.1	Introduction	161
3.2	Bayes Procedures	161
3.3	Minimax Procedures	170
3.4	Unbiased Estimation and Risk Inequalities	176
3.4.1	Unbiased Estimation, Survey Sampling	176
3.4.2	The Information Inequality	179
3.5	Nondecision Theoretic Criteria	188
3.5.1	Computation	188
3.5.2	Interpretability	189
3.5.3	Robustness	190
3.6	Problems and Complements	197
3.7	Notes	210
3.8	References	211
<b>4</b>	<b>TESTING AND CONFIDENCE REGIONS</b>	<b>213</b>
4.1	Introduction	213
4.2	Choosing a Test Statistic: The Neyman-Pearson Lemma	223
4.3	Uniformly Most Powerful Tests and Monotone Likelihood Ratio Models	227
4.4	Confidence Bounds, Intervals, and Regions	233

4.5	The Duality Between Confidence Regions and Tests	241
4.6	Uniformly Most Accurate Confidence Bounds	248
4.7	Frequentist and Bayesian Formulations	251
4.8	Prediction Intervals	252
4.9	Likelihood Ratio Procedures	255
4.9.1	Introduction	255
4.9.2	Tests for the Mean of a Normal Distribution-Matched Pair Experiments	257
4.9.3	Tests and Confidence Intervals for the Difference in Means of Two Normal Populations	261
4.9.4	The Two-Sample Problem with Unequal Variances	264
4.9.5	Likelihood Ratio Procedures for Bivariate Normal Distributions	266
4.10	Problems and Complements	269
4.11	Notes	295
4.12	References	295
<b>5</b>	<b>ASYMPTOTIC APPROXIMATIONS</b>	<b>297</b>
5.1	Introduction: The Meaning and Uses of Asymptotics	297
5.2	Consistency	301
5.2.1	Plug-In Estimates and MLEs in Exponential Family Models	301
5.2.2	Consistency of Minimum Contrast Estimates	304
5.3	First- and Higher-Order Asymptotics: The Delta Method with Applications	306
5.3.1	The Delta Method for Moments	306
5.3.2	The Delta Method for In Law Approximations	311
5.3.3	Asymptotic Normality of the Maximum Likelihood Estimate in Exponential Families	322
5.4	Asymptotic Theory in One Dimension	324
5.4.1	Estimation: The Multinomial Case	324
5.4.2	Asymptotic Normality of Minimum Contrast and $M$ -Estimates	327
5.4.3	Asymptotic Normality and Efficiency of the MLE	331
5.4.4	Testing	332
5.4.5	Confidence Bounds	336
5.5	Asymptotic Behavior and Optimality of the Posterior Distribution	337
5.6	Problems and Complements	345
5.7	Notes	362
5.8	References	363

<b>6</b>	<b>INFERENCE IN THE MULTIPARAMETER CASE</b>	<b>365</b>
6.1	Inference for Gaussian Linear Models	365
6.1.1	The Classical Gaussian Linear Model	366
6.1.2	Estimation	369
6.1.3	Tests and Confidence Intervals	374
6.2	Asymptotic Estimation Theory in $p$ Dimensions	383
6.2.1	Estimating Equations	384
6.2.2	Asymptotic Normality and Efficiency of the MLE	386
6.2.3	The Posterior Distribution in the Multiparameter Case	391
6.3	Large Sample Tests and Confidence Regions	392
6.3.1	Asymptotic Approximation to the Distribution of the Likelihood Ratio Statistic	392
6.3.2	Wald's and Rao's Large Sample Tests	398
6.4	Large Sample Methods for Discrete Data	400
6.4.1	Goodness-of-Fit in a Multinomial Model. Pearson's $\chi^2$ Test	401
6.4.2	Goodness-of-Fit to Composite Multinomial Models. Contingency Tables	403
6.4.3	Logistic Regression for Binary Responses	408
6.5	Generalized Linear Models	411
6.6	Robustness Properties and Semiparametric Models	417
6.7	Problems and Complements	422
6.8	Notes	438
6.9	References	438
<b>A</b>	<b>A REVIEW OF BASIC PROBABILITY THEORY</b>	<b>441</b>
A.1	The Basic Model	441
A.2	Elementary Properties of Probability Models	443
A.3	Discrete Probability Models	443
A.4	Conditional Probability and Independence	444
A.5	Compound Experiments	446
A.6	Bernoulli and Multinomial Trials, Sampling With and Without Replacement	447
A.7	Probabilities on Euclidean Space	448
A.8	Random Variables and Vectors: Transformations	451
A.9	Independence of Random Variables and Vectors	453
A.10	The Expectation of a Random Variable	454
A.11	Moments	456
A.12	Moment and Cumulant Generating Functions	459



A.13 Some Classical Discrete and Continuous Distributions	460
A.14 Modes of Convergence of Random Variables and Limit Theorems	466
A.15 Further Limit Theorems and Inequalities	468
A.16 Poisson Process	472
A.17 Notes	474
A.18 References	475
<b>B ADDITIONAL TOPICS IN PROBABILITY AND ANALYSIS</b>	<b>477</b>
B.1 Conditioning by a Random Variable or Vector	477
B.1.1 The Discrete Case	477
B.1.2 Conditional Expectation for Discrete Variables	479
B.1.3 Properties of Conditional Expected Values	480
B.1.4 Continuous Variables	482
B.1.5 Comments on the General Case	484
B.2 Distribution Theory for Transformations of Random Vectors	485
B.2.1 The Basic Framework	485
B.2.2 The Gamma and Beta Distributions	488
B.3 Distribution Theory for Samples from a Normal Population	491
B.3.1 The $\chi^2$ , $F$ , and $t$ Distributions	491
B.3.2 Orthogonal Transformations	494
B.4 The Bivariate Normal Distribution	497
B.5 Moments of Random Vectors and Matrices	502
B.5.1 Basic Properties of Expectations	502
B.5.2 Properties of Variance	503
B.6 The Multivariate Normal Distribution	506
B.6.1 Definition and Density	506
B.6.2 Basic Properties. Conditional Distributions	508
B.7 Convergence for Random Vectors: $O_P$ and $o_P$ Notation	511
B.8 Multivariate Calculus	516
B.9 Convexity and Inequalities	518
B.10 Topics in Matrix Theory and Elementary Hilbert Space Theory	519
B.10.1 Symmetric Matrices	519
B.10.2 Order on Symmetric Matrices	520
B.10.3 Elementary Hilbert Space Theory	521
B.11 Problems and Complements	524
B.12 Notes	538
B.13 References	539

<b>C TABLES</b>	<b>541</b>
Table I The Standard Normal Distribution	542
Table I' Auxilliary Table of the Standard Normal Distribution	543
Table II $t$ Distribution Critical Values	544
Table III $\chi^2$ Distribution Critical Values	545
Table IV $F$ Distribution Critical Values	546
<b>INDEX</b>	<b>547</b>

---

---

# VOLUME II CONTENTS

<b>I</b>	<b>INTRODUCTION TO VOLUME II</b>	<b>1</b>
I.1	Tests of Goodness of Fit and the Brownian Bridge	5
I.2	Testing Goodness of Fit to Parametric Hypotheses	5
I.3	Regular Parameters. Minimum Distance Estimates	6
I.4	Permutation Tests	8
I.5	Estimation of Irregular Parameters	8
I.6	Stein and Empirical Bayes Estimation	10
I.7	Model Selection	11
I.8	Problems and Complements	15
I.9	Notes	20
<b>7</b>	<b>TOOLS FOR ASYMPTOTIC ANALYSIS</b>	<b>21</b>
7.1	Weak Convergence in Function Spaces	21
7.1.1	Stochastic Processes and Weak Convergence	21
7.1.2	Maximal Inequalities	28
7.1.3	Empirical Processes on Function Spaces	31
7.2	The Delta Method in Infinite Dimensional Space	38
7.2.1	Influence Functions and the Gâteaux Derivative	38
7.2.2	The Quantile Process	47
7.3	Further Expansions	51
7.3.1	The von Mises Expansion	51
7.3.2	The Hoeffding/Analysis of Variance Expansion	54
7.4	Problems and Complements	62
7.5	Notes	71

<b>8</b>	<b>DISTRIBUTION-FREE, UNBIASED AND EQUIVARIANT PROCEDURES</b>	<b>72</b>
8.1	Introduction	72
8.2	Similarity and Completeness	73
8.2.1	Testing	73
8.2.2	Testing Optimality Theory	83
8.2.3	Estimation	86
8.3	Invariance, Equivariance and Minimax Procedures	91
8.3.1	Group Models	91
8.3.2	Group Models and Decision Theory	93
8.3.3	Characterizing Invariant Tests	95
8.3.4	Characterizing Equivariant Estimates	101
8.3.5	Minimaxity for Tests: Application to Group Models	102
8.3.6	Minimax Estimation, Admissibility, and Steinian Shrinkage	106
8.4	Problems and Complements	111
8.5	Notes	122
<b>9</b>	<b>INFERENCE IN SEMIPARAMETRIC MODELS</b>	<b>123</b>
9.1	ESTIMATION IN SEMIPARAMETRIC MODELS	123
9.1.1	Selected Examples	123
9.1.2	Regularization. Modified Maximum Likelihood	131
9.1.3	Other Modified and Approximate Likelihoods	140
9.1.4	Sieves and Regularization	143
9.2	Asymptotics. Consistency and Asymptotic Normality	149
9.2.1	A General Consistency Criterion	149
9.2.2	Asymptotics for Selected Models	151
9.3	Efficiency in Semiparametric Models	159
9.4	Tests and Empirical Process Theory	172
9.5	Asymptotic Properties of Likelihoods. Contiguity	177
9.6	Problems and Complements	189
9.7	Notes	205
<b>10</b>	<b>MONTE CARLO METHODS</b>	<b>207</b>
10.1	The Nature of Monte Carlo Methods	207
10.2	Three Basic Monte Carlo Methods	210
10.2.1	Simple Monte Carlo	211
10.2.2	Importance Sampling	212
10.2.3	Rejective Sampling	213

10.3	The Bootstrap	215
10.3.1	Bootstrap Samples and Bias Corrections	216
10.3.2	Bootstrap Variance and Confidence Bounds	220
10.3.3	The General i.i.d. Nonparametric Bootstrap	222
10.3.4	Asymptotic Theory for the Bootstrap	225
10.3.5	Examples where Efron's Bootstrap Fails. The $m$ out of $n$ Bootstraps	230
10.4	Markov Chain Monte Carlo	232
10.4.1	The Basic MCMC Framework	232
10.4.2	Metropolis Sampling Algorithms	233
10.4.3	The Gibbs Samplers	237
10.4.4	Speed of Convergence of MCMC	241
10.5	Applications of MCMC to Bayesian and Frequentist Inference	243
10.6	Problems and Complements	250
10.7	Notes	256
<b>11</b>	<b>NONPARAMETRIC INFERENCE FOR FUNCTIONS OF ONE VARIABLE</b>	<b>257</b>
11.1	Introduction	257
11.2	Convolution Kernel Estimates on $R$	258
11.2.1	Uniform Local Behavior of Kernel Density Estimates	261
11.2.2	Global Behavior of Convolution Kernel Estimates	263
11.2.3	Performance and Bandwidth Choice	264
11.2.4	Discussion of convolution kernel estimates	265
11.3	Minimum Contrast Estimates: Reducing Boundary Bias	266
11.4	Regularization and Nonlinear Density Estimates	272
11.4.1	Regularization and Roughness Penalties	272
11.4.2	Sieves. Machine Learning. Log Density Estimation	273
11.4.3	Nearest Neighbour Density Estimates	276
11.5	Confidence Regions	277
11.6	Nonparametric Regression for one Covariate	279
11.6.1	Estimation Principles	279
11.6.2	Asymptotic Bias and Variance Calculations	282
11.7	Problems and Complements	289
<b>12</b>	<b>PREDICTION AND MACHINE LEARNING</b>	<b>299</b>
12.1	Introduction	299
12.1.1	Statistical Approaches to Modeling and Analyzing Multidimensional data. Sieves	301

12.1.2	Machine Learning Approaches	305
12.1.3	Outline	307
12.2	Classification and Prediction	307
12.2.1	Multivariate Density and Regression Estimation	307
12.2.2	Bayes Rule and Nonparametric Classification	312
12.2.3	Sieve Methods	314
12.2.4	Machine Learning Approaches	316
12.3	Asymptotics	324
12.3.1	Optimal Prediction in Parametric Regression Models	326
12.3.2	Optimal Rates of Convergence for Estimation and Prediction in Nonparametric Models	329
12.3.3	The Gaussian White Noise (GWN) Model	338
12.3.4	Minmax Bounds on IMSE for Subsets of the GWN Model	340
12.3.5	Sparse Submodels	342
12.4	Oracle Inequalities	344
12.4.1	Stein's Unbiased Risk Estimate	346
12.4.2	Oracle Inequality for Shrinkage Estimators	347
12.4.3	Oracle Inequality and Adaptive Minimax Rate for Truncated Estimates	348
12.4.4	An Oracle Inequality for Classification	350
12.5	Performance and Tuning via Cross Validation	353
12.5.1	Cross Validation for Tuning Parameter Choice	354
12.5.2	Cross Validation for Measuring Performance	358
12.6	Model Selection and Dimension Reduction	359
12.6.1	A Bayesian Criterion for Model Selection	360
12.6.2	Inference after Model Selection	364
12.6.3	Dimension Reduction via Principal Component Analysis	366
12.7	Topics Untouched and Current Frontiers	367
12.8	Problems and Complements	371
<b>D</b>	<b>APPENDIX D. SUPPLEMENTS TO TEXT</b>	<b>385</b>
D.1	Probability Results	385
D.2	Supplements to Section 7.1	387
D.3	Supplement to Section 7.2	390
D.4	Supplement to Section 9.2.2	391
D.5	Supplement to Section 10.4	392
D.6	Supplement to Section 11.6	397

---

D.7 Supplement to Section 12.2.2	399
D.8 Problems and Complements	405
<b>E SOLUTIONS FOR VOL. II</b>	<b>410</b>
<b>REFERENCES</b>	<b>423</b>
<b>SUBJECT INDEX</b>	<b>438</b>

This page intentionally left blank