

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)

Large-Scale Inference

We live in a new age for statistical inference, where modern scientific technology such as microarrays and fMRI machines routinely produce thousands and sometimes millions of parallel data sets, each with its own estimation or testing problem. Doing thousands of problems at once involves more than repeated application of classical methods.

Taking an empirical Bayes approach, Bradley Efron, inventor of the bootstrap, shows how information accrues across problems in a way that combines Bayesian and frequentist ideas. Estimation, testing, and prediction blend in this framework, producing opportunities for new methodologies of increased power. New difficulties also arise, easily leading to flawed inferences. This book takes a careful look at both the promise and pitfalls of large-scale statistical inference, with particular attention to false discovery rates, the most successful of the new statistical techniques. Emphasis is on the inferential ideas underlying technical developments, illustrated using a large number of real examples.

BRADLEY EFRON is Max H. Stein Professor of Statistics and Biostatistics at the Stanford University School of Humanities and Sciences, and the Department of Health Research and Policy at the School of Medicine.

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)

INSTITUTE OF MATHEMATICAL STATISTICS
MONOGRAPHS

Editorial Board

D. R. Cox (University of Oxford)

B. Hambly (University of Oxford)

S. Holmes (Stanford University)

X.-L. Meng (Harvard University)

IMS Monographs are concise research monographs of high quality on any branch of statistics or probability of sufficient interest to warrant publication as books. Some concern relatively traditional topics in need of up-to-date assessment. Others are on emerging themes. In all cases the objective is to provide a balanced view of the field.

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)

Large-Scale Inference
Empirical Bayes Methods for
Estimation, Testing, and Prediction

BRADLEY EFRON

Stanford University



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9781107619678

© Bradley Efron 2010

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2010
First paperback edition 2013

Printed and bound in the United Kingdom by the MPG Books Group

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-19249-1 Hardback
ISBN 978-1-107-61967-8 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)

Contents

<i>Prologue</i>	ix
<i>Acknowledgments</i>	xii
1 Empirical Bayes and the James–Stein Estimator	1
1.1 Bayes Rule and Multivariate Normal Estimation	2
1.2 Empirical Bayes Estimation	4
1.3 Estimating the Individual Components	7
1.4 Learning from the Experience of Others	10
1.5 Empirical Bayes Confidence Intervals	12
Notes	14
2 Large-Scale Hypothesis Testing	15
2.1 A Microarray Example	15
2.2 Bayesian Approach	17
2.3 Empirical Bayes Estimates	20
2.4 $\overline{\text{Fdr}}(\mathcal{Z})$ as a Point Estimate	22
2.5 Independence versus Correlation	26
2.6 Learning from the Experience of Others II	27
Notes	28
3 Significance Testing Algorithms	30
3.1 p -Values and z -Values	31
3.2 Adjusted p -Values and the FWER	34
3.3 Stepwise Algorithms	37
3.4 Permutation Algorithms	39
3.5 Other Control Criteria	43
Notes	45
4 False Discovery Rate Control	46
4.1 True and False Discoveries	46
4.2 Benjamini and Hochberg’s FDR Control Algorithm	48
4.3 Empirical Bayes Interpretation	52

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)

vi	<i>Contents</i>	
4.4	Is FDR Control “Hypothesis Testing”?	58
4.5	Variations on the Benjamini–Hochberg Algorithm	59
4.6	Fdr and Simultaneous Tests of Correlation	64
	Notes	69
5	Local False Discovery Rates	70
5.1	Estimating the Local False Discovery Rate	70
5.2	Poisson Regression Estimates for $f(z)$	74
5.3	Inference and Local False Discovery Rates	77
5.4	Power Diagnostics	83
	Notes	88
6	Theoretical, Permutation, and Empirical Null Distributions	89
6.1	Four Examples	90
6.2	Empirical Null Estimation	97
6.3	The MLE Method for Empirical Null Estimation	102
6.4	Why the Theoretical Null May Fail	105
6.5	Permutation Null Distributions	109
	Notes	112
7	Estimation Accuracy	113
7.1	Exact Covariance Formulas	115
7.2	Rms Approximations	121
7.3	Accuracy Calculations for General Statistics	126
7.4	The Non-Null Distribution of z -Values	132
7.5	Bootstrap Methods	138
	Notes	139
8	Correlation Questions	141
8.1	Row and Column Correlations	141
8.2	Estimating the Root Mean Square Correlation	145
8.3	Are a Set of Microarrays Independent of Each Other?	149
8.4	Multivariate Normal Calculations	153
8.5	Count Correlations	159
	Notes	162
9	Sets of Cases (Enrichment)	163
9.1	Randomization and Permutation	164
9.2	Efficient Choice of a Scoring Function	170
9.3	A Correlation Model	174
9.4	Local Averaging	181
	Notes	184

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)*Contents*

vii

10	Combination, Relevance, and Comparability	185
10.1	The Multi-Class Model	187
10.2	Small Subclasses and Enrichment	192
10.3	Relevance	196
10.4	Are Separate Analyses Legitimate?	199
10.5	Comparability	206
	Notes	209
11	Prediction and Effect Size Estimation	211
11.1	A Simple Model	213
11.2	Bayes and Empirical Bayes Prediction Rules	217
11.3	Prediction and Local False Discovery Rates	223
11.4	Effect Size Estimation	227
11.5	The Missing Species Problem	233
	Notes	240
<i>Appendix A</i>	Exponential Families	243
A.1	Multiparameter Exponential Families	245
A.2	Lindsey's Method	247
<i>Appendix B</i>	Data Sets and Programs	249
	<i>References</i>	251
	<i>Index</i>	258

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)

Prologue

At the risk of drastic oversimplification, the history of statistics as a recognized discipline can be divided into three eras:

- 1 The age of Quetelet and his successors, in which huge census-level data sets were brought to bear on simple but important questions: Are there more male than female births? Is the rate of insanity rising?
- 2 The classical period of Pearson, Fisher, Neyman, Hotelling, and their successors, intellectual giants who developed a theory of optimal inference capable of wringing every drop of information out of a scientific experiment. The questions dealt with still tended to be simple — Is treatment A better than treatment B? — but the new methods were suited to the kinds of small data sets individual scientists might collect.
- 3 The era of scientific mass production, in which new technologies typified by the microarray allow a single team of scientists to produce data sets of a size Quetelet would envy. But now the flood of data is accompanied by a deluge of questions, perhaps thousands of estimates or hypothesis tests that the statistician is charged with answering together; not at all what the classical masters had in mind.

The response to this onslaught of data has been a tremendous burst of statistical methodology, impressively creative, showing an attractive ability to come to grips with changed circumstances, and at the same time highly speculative. There is plenty of methodology in what follows, but that is not the main theme of the book. My primary goal has been to ground the methodology in familiar principles of statistical inference.

This is where the “empirical Bayes” in my subtitle comes into consideration. By their nature, empirical Bayes arguments combine frequentist and Bayesian elements in analyzing problems of repeated structure. Repeated structures are just what scientific mass production excels at, e.g., expression levels comparing sick and healthy subjects for thousands of genes at the same time by means of microarrays. At their best, the new methodolo-

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)

gies are successful from both Bayes and frequentist viewpoints, which is what my empirical Bayes arguments are intended to show.

False discovery rates, Benjamini and Hochberg's seminal contribution, is the great success story of the new methodology. Much of what follows is an attempt to explain that success in empirical Bayes terms. FDR, indeed, has strong credentials in both the Bayesian and frequentist camps, always a good sign that we are on the right track, as well as a suggestion of fruitful empirical Bayes explication.

The later chapters are at pains to show the limitations of current large-scale statistical practice: Which cases should be combined in a single analysis? How do we account for notions of relevance between cases? What is the correct null hypothesis? How do we handle correlations? Some helpful theory is provided in answer, but much of the argumentation is by example, with graphs and figures playing a major role. The examples are real ones, collected in a sometimes humbling decade of large-scale data analysis at the Stanford School of Medicine and Department of Statistics. (My examples here are mainly biomedical, but of course that has nothing to do with the basic ideas, which are presented with no prior medical or biological knowledge assumed.)

In moving beyond the confines of classical statistics, we are also moving outside its wall of protection. Fisher, Neyman et al. fashioned an almost perfect inferential machine for small-scale estimation and testing problems. It is hard to go wrong using maximum likelihood estimation or a t -test on a typical small data set. I have found it very easy to go wrong with huge data sets and thousands of questions to answer at once. Without claiming a cure, I hope the various examples at least help identify the symptoms.

The classical era of statistics can itself be divided into two periods: the first half of the 20th century, during which basic theory was developed, and then a great methodological expansion of that theory in the second half. Empirical Bayes stands as a striking exception. Emerging in the 1950s in two branches identified with Charles Stein and Herbert Robbins, it represented a genuinely new initiative in statistical theory. The Stein branch concerned normal estimation theory, while the Robbins branch was more general, being applicable to both estimation and hypothesis testing.

Typical large-scale applications have been more concerned with testing than estimation. If judged by chapter titles, the book seems to share this imbalance, but that is misleading. Empirical Bayes blurs the line between testing and estimation as well as between frequentism and Bayesianism. Much of what follows is an attempt to say how well we can estimate a testing procedure, for example how accurately can a null distribution be esti-

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)

mated? The false discovery rate procedure itself strays far from the spirit of classical hypothesis testing, as discussed in Chapter 4.

About this book: it is written for readers with at least a second course in statistics as background. The mathematical level is not daunting — mainly multidimensional calculus, probability theory, and linear algebra — though certain parts are more intricate, particularly in Chapters 3 and 7 (which can be scanned or skipped at first reading). There are almost no asymptotics. Exercises are interspersed in the text as they arise (rather than being lumped together at the end of chapters), where they mostly take the place of statements like “It is easy to see ...” or “It can be shown ...”. Citations are concentrated in the **Notes** section at the end of each chapter. There are two brief appendices, one listing basic facts about exponential families, the second concerning access to some of the programs and data sets featured in the text.

I have perhaps abused the “mono” in monograph by featuring methods from my own work of the past decade. This is not a survey or a textbook, though I hope it can be used for a graduate-level lecture course. In fact, I am not trying to sell any particular methodology, my main interest as stated above being how the methods mesh with basic statistical theory.

There are at least three excellent books for readers who wish to see different points of view. Working backwards in time, Dudoit and van der Laan’s 2009 *Multiple Testing Procedures with Applications to Genomics* emphasizes the control of Type I error. It is a successor to *Resampling-based Multiple Testing: Examples and Methods for p-Value Adjustment* (Westfall and Young, 1993), which now looks far ahead of its time. Miller’s classic text, *Simultaneous Statistical Inference* (1981), beautifully describes the development of multiple testing before the era of large-scale data sets, when “multiple” meant somewhere between two and ten problems, not thousands.

I chose the adjective *large-scale* to describe massive data analysis problems rather than “multiple,” “high-dimensional,” or “simultaneous,” because of its bland neutrality with regard to estimation, testing, or prediction, as well as its lack of identification with specific methodologies. My intention is not to have the last word here, and in fact I hope for and expect a healthy development of new ideas in dealing with the burgeoning statistical problems of the 21st century.

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Frontmatter

[More information](#)

Acknowledgments

The Institute of Mathematical Statistics has begun an ambitious new monograph series in statistics, and I am grateful to the editors David Cox, Xiao-Li Meng, and Susan Holmes for their encouragement, and for letting me in on the ground floor. Diana Gillooly, the editor at Cambridge University Press (now in its fifth century!) has been supportive, encouraging, and gentle in correcting my literary abuses. My colleague Elizabeth Halloran has shown a sharp eye for faulty derivations and confused wording. Many of my Stanford colleagues and students have helped greatly in the book's final development, with Rob Tibshirani and Omkar Muralidharan deserving special thanks. Most of all, I am grateful to my associate Cindy Kirby for her tireless work in transforming my handwritten pages into the book you see here.

Bradley Efron
Department of Statistics
Stanford University