# Contents

*Preface* xv  
*Acknowledgments* xviii  
*Notation* xix  

**Part I Classic Statistical Inference** 1  

1 **Algorithms and Inference** 3  
1.1 A Regression Example 4  
1.2 Hypothesis Testing 8  
1.3 Notes 11  

2 **Frequentist Inference** 12  
2.1 Frequentism in Practice 14  
2.2 Frequentist Optimality 18  
2.3 Notes and Details 20  

3 **Bayesian Inference** 22  
3.1 Two Examples 24  
3.2 Uninformative Prior Distributions 28  
3.3 Flaws in Frequentist Inference 30  
3.4 A Bayesian/Frequentist Comparison List 33  
3.5 Notes and Details 36  

4 **Fisherian Inference and Maximum Likelihood Estimation** 38  
4.1 Likelihood and Maximum Likelihood 38  
4.2 Fisher Information and the MLE 41  
4.3 Conditional Inference 45  
4.4 Permutation and Randomization 49  
4.5 Notes and Details 51  

5 **Parametric Models and Exponential Families** 53

ix
x Contents

5.1 Univariate Families 54
5.2 The Multivariate Normal Distribution 55
5.3 Fisher’s Information Bound for Multiparameter Families 59
5.4 The Multinomial Distribution 61
5.5 Exponential Families 64
5.6 Notes and Details 69

Part II Early Computer-Age Methods 73

6 Empirical Bayes 75
6.1 Robbins’ Formula 75
6.2 The Missing-Species Problem 78
6.3 A Medical Example 84
6.4 Indirect Evidence 1 88
6.5 Notes and Details 88

7 James–Stein Estimation and Ridge Regression 91
7.1 The James–Stein Estimator 91
7.2 The Baseball Players 94
7.3 Ridge Regression 97
7.4 Indirect Evidence 2 102
7.5 Notes and Details 104

8 Generalized Linear Models and Regression Trees 108
8.1 Logistic Regression 109
8.2 Generalized Linear Models 116
8.3 Poisson Regression 120
8.4 Regression Trees 124
8.5 Notes and Details 128

9 Survival Analysis and the EM Algorithm 131
9.1 Life Tables and Hazard Rates 131
9.2 Censored Data and the Kaplan–Meier Estimate 134
9.3 The Log-Rank Test 139
9.4 The Proportional Hazards Model 143
9.5 Missing Data and the EM Algorithm 146
9.6 Notes and Details 150

10 The Jackknife and the Bootstrap 155
10.1 The Jackknife Estimate of Standard Error 156
10.2 The Nonparametric Bootstrap 159
10.3 Resampling Plans 162
## Contents

10.4 The Parametric Bootstrap 169  
10.5 Influence Functions and Robust Estimation 174  
10.6 Notes and Details 177  

11 Bootstrap Confidence Intervals 181  
11.1 Neyman’s Construction for One-Parameter Problems 181  
11.2 The Percentile Method 185  
11.3 Bias-Corrected Confidence Intervals 190  
11.4 Second-Order Accuracy 192  
11.5 Bootstrap-t Intervals 195  
11.6 Objective Bayes Intervals and the Confidence Distribution 198  
11.7 Notes and Details 204  

12 Cross-Validation and $C_p$ Estimates of Prediction Error 208  
12.1 Prediction Rules 208  
12.2 Cross-Validation 213  
12.3 Covariance Penalties 218  
12.4 Training, Validation, and Ephemeral Predictors 227  
12.5 Notes and Details 230  

13 Objective Bayes Inference and MCMC 233  
13.1 Objective Prior Distributions 234  
13.2 Conjugate Prior Distributions 237  
13.3 Model Selection and the Bayesian Information Criterion 243  
13.4 Gibbs Sampling and MCMC 251  
13.5 Example: Modeling Population Admixture 256  
13.6 Notes and Details 261  

14 Postwar Statistical Inference and Methodology 264  

### Part III Twenty-First-Century Topics 269

15 Large-Scale Hypothesis Testing and FDRs 271  
15.1 Large-Scale Testing 272  
15.2 False-Discovery Rates 275  
15.3 Empirical Bayes Large-Scale Testing 278  
15.4 Local False-Discovery Rates 282  
15.5 Choice of the Null Distribution 286  
15.6 Relevance 290  
15.7 Notes and Details 294  

16 Sparse Modeling and the Lasso 298
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.1</td>
<td>Forward Stepwise Regression</td>
<td>299</td>
</tr>
<tr>
<td>16.2</td>
<td>The Lasso</td>
<td>303</td>
</tr>
<tr>
<td>16.3</td>
<td>Fitting Lasso Models</td>
<td>308</td>
</tr>
<tr>
<td>16.4</td>
<td>Least-Angle Regression</td>
<td>309</td>
</tr>
<tr>
<td>16.5</td>
<td>Fitting Generalized Lasso Models</td>
<td>313</td>
</tr>
<tr>
<td>16.6</td>
<td>Post-Selection Inference for the Lasso</td>
<td>317</td>
</tr>
<tr>
<td>16.7</td>
<td>Connections and Extensions</td>
<td>319</td>
</tr>
<tr>
<td>16.8</td>
<td>Notes and Details</td>
<td>321</td>
</tr>
<tr>
<td>17.1</td>
<td>Random Forests</td>
<td>324</td>
</tr>
<tr>
<td>17.2</td>
<td>Boosting with Squared-Error Loss</td>
<td>325</td>
</tr>
<tr>
<td>17.3</td>
<td>Gradient Boosting</td>
<td>333</td>
</tr>
<tr>
<td>17.4</td>
<td>Adaboost: the Original Boosting Algorithm</td>
<td>338</td>
</tr>
<tr>
<td>17.5</td>
<td>Connections and Extensions</td>
<td>341</td>
</tr>
<tr>
<td>17.6</td>
<td>Notes and Details</td>
<td>345</td>
</tr>
<tr>
<td>18.1</td>
<td>Neural Networks and the Handwritten Digit Problem</td>
<td>351</td>
</tr>
<tr>
<td>18.2</td>
<td>Fitting a Neural Network</td>
<td>353</td>
</tr>
<tr>
<td>18.3</td>
<td>Autoencoders</td>
<td>356</td>
</tr>
<tr>
<td>18.4</td>
<td>Deep Learning</td>
<td>362</td>
</tr>
<tr>
<td>18.5</td>
<td>Learning a Deep Network</td>
<td>364</td>
</tr>
<tr>
<td>18.6</td>
<td>Notes and Details</td>
<td>368</td>
</tr>
<tr>
<td>19.1</td>
<td>Support-Vector Machines and Kernel Methods</td>
<td>375</td>
</tr>
<tr>
<td>19.2</td>
<td>Optimal Separating Hyperplane</td>
<td>376</td>
</tr>
<tr>
<td>19.3</td>
<td>Soft-Margin Classifier</td>
<td>378</td>
</tr>
<tr>
<td>19.4</td>
<td>SVM Criterion as Loss Plus Penalty</td>
<td>379</td>
</tr>
<tr>
<td>19.5</td>
<td>Computations and the Kernel Trick</td>
<td>381</td>
</tr>
<tr>
<td>19.6</td>
<td>Function Fitting Using Kernels</td>
<td>384</td>
</tr>
<tr>
<td>19.7</td>
<td>Example: String Kernels for Protein Classification</td>
<td>385</td>
</tr>
<tr>
<td>19.8</td>
<td>SVMs: Concluding Remarks</td>
<td>387</td>
</tr>
<tr>
<td>19.9</td>
<td>Kernel Smoothing and Local Regression</td>
<td>387</td>
</tr>
<tr>
<td>20.1</td>
<td>Inference After Model Selection</td>
<td>390</td>
</tr>
<tr>
<td>20.2</td>
<td>Combined Bayes–Frequentist Estimation</td>
<td>394</td>
</tr>
<tr>
<td>20.3</td>
<td>Simultaneous Confidence Intervals</td>
<td>395</td>
</tr>
<tr>
<td>20.4</td>
<td>Accuracy After Model Selection</td>
<td>402</td>
</tr>
<tr>
<td>20.5</td>
<td>Selection Bias</td>
<td>408</td>
</tr>
<tr>
<td>20.6</td>
<td>Notes and Details</td>
<td>412</td>
</tr>
</tbody>
</table>
21 Empirical Bayes Estimation Strategies 421
   21.1 Bayes Deconvolution 421
   21.2 g-Modeling and Estimation 424
   21.3 Likelihood, Regularization, and Accuracy 427
   21.4 Two Examples 432
   21.5 Generalized Linear Mixed Models 437
   21.6 Deconvolution and f-Modeling 440
   21.7 Notes and Details 444

Epilogue 446
References 453
Author Index 463
Subject Index 467