

**Larry Wasserman**

# **All of Nonparametric Statistics**

**A Concise Course in Nonparametric Statistical  
Inference**

Springer  
Berlin Heidelberg New York  
Barcelona Hong Kong  
London Milan Paris  
Tokyo



# Preface

There are many books on various aspects of nonparametric inference such as density estimation, nonparametric regression, bootstrapping, and wavelets methods. But it is hard to find all these topics covered in one place. The goal of this text is to provide the reader with a single book where they can find a brief account of many of the modern topics in nonparametric inference.

The book is aimed at Master's level or Ph.D. level statistics and computer science students. It is also suitable for researchers in statistics, machine learning and data mining who want to get up to speed quickly on modern nonparametric methods. My goal is to quickly acquaint the reader with the basic concepts in many areas rather than tackling any one topic in great detail. In the interest of covering a wide range of topics, while keeping the book short, I have opted to omit most proofs. Bibliographic remarks point the reader to references which contain further details. Of course, I have had to choose topics to include and to omit, the title notwithstanding. For the most part, I decided to omit topics that are too big to cover in one chapter. For example, I do not cover classification or nonparametric Bayesian inference.

The book developed from my lecture notes for a half-semester (twenty hours) course populated mainly by Masters' level students. For Ph.D. level students, the instructor may want to cover some of the material in more depth and require the students to fill in proofs of some of the theorems. Throughout I have attempted to follow one basic principle: never give an estimator without giving a confidence set.

The book has a mixture of methods and theory. The material is meant to complement more purely method oriented texts such as Hastie et al. (2001) and Ruppert et al. (2003).

After the Introduction in Chapter one, Chapters two and three cover inference related to the empirical CDF such as the nonparametric delta method and the bootstrap. Chapter four to six cover basic smoothing methods. Chapters seven to nine have a higher theoretical content and are more demanding. The theory in Chapter seven lays the foundation for the orthogonal function methods in Chapters eight and nine. Chapter ten surveys some of the omitted topics.

I assume that the student has had a course on mathematical statistics such as Casella and Berger (2002) or Wasserman (2004). In particular, I assume that the following concepts are familiar to the reader: distribution functions, convergence in probability, convergence in distribution, almost sure convergence, likelihood functions, maximum likelihood, confidence intervals, the delta method, bias, mean squared error, and Bayes estimators. These background concepts are reviewed briefly in Chapter 1.

Data sets and code can be found at:

`www.stat.cmu.edu/~larry/all-of-nonpar`

I need to make some disclaimers. First, the topics in this book fall under the rubric of “modern nonparametrics.” The omission of traditional methods like rank tests and so on is not intended to belittle their importance. Second, I make heavy use of large sample methods. This is partly because I think that statistics is, largely, most successful and useful in large sample situations, and partly because it is often easier to construct large sample, nonparametric methods. The reader should be aware that large sample methods can of course go awry when used without appropriate caution.

I would like to thank the following people for providing feedback and suggestions: Larry Brown, Taeryon Choi, Ed George, Nils Hjort, Woncheol Jang, Feng Liang, Catherine Loader, Jiayang Sun, Rob Tibshirani and special thanks to the reviewers: David Ruppert, Javier Rojo, Chris Jones, David W. Scott and one anonymous reader, for carefully reading the book and making suggesting numerous improvements. Thanks also go to my colleague Chris Genovese for lots of advice and for writing the  $\LaTeX$  macros for the layout of the book. I am indebted to John Kimmel who has been supportive and helpful and did not rebel against the crazy title. Finally, thanks to my wife Isabella Verdinelli for her love and support.

*Larry Wasserman*  
Pittsburgh, Pennsylvania  
March 2005

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                                   | <b>1</b>  |
| 1.1      | What is Nonparametric Inference? . . . . .            | 1         |
| 1.2      | Notation and Background . . . . .                     | 2         |
| 1.3      | Confidence Sets . . . . .                             | 5         |
| 1.4      | Useful Inequalities . . . . .                         | 8         |
| 1.5      | Bibliographic Remarks . . . . .                       | 10        |
| 1.6      | Exercises . . . . .                                   | 10        |
| <b>2</b> | <b>Estimating the CDF and Statistical Functionals</b> | <b>13</b> |
| 2.1      | The CDF . . . . .                                     | 13        |
| 2.2      | Estimating Statistical Functionals . . . . .          | 15        |
| 2.3      | Influence Functions . . . . .                         | 18        |
| 2.4      | Empirical Probability Distributions . . . . .         | 21        |
| 2.5      | Bibliographic Remarks . . . . .                       | 23        |
| 2.6      | Appendix . . . . .                                    | 23        |
| 2.7      | Exercises . . . . .                                   | 24        |
| <b>3</b> | <b>The Bootstrap and The Jackknife</b>                | <b>27</b> |
| 3.1      | The Jackknife . . . . .                               | 27        |
| 3.2      | The Bootstrap . . . . .                               | 30        |
| 3.3      | Parametric Bootstrap . . . . .                        | 31        |
| 3.4      | Bootstrap Confidence Intervals . . . . .              | 32        |
| 3.5      | Some Theory . . . . .                                 | 35        |
| 3.6      | Bibliographic Remarks . . . . .                       | 37        |

|          |  |            |
|----------|--|------------|
| 3.7      | Appendix . . . . .   | 37         |
| 3.8      | Exercises . . . . .  | 39         |
| <b>4</b> | <b>Smoothing: General Concepts</b>                         | <b>43</b>  |
| 4.1      | The Bias-Variance Tradeoff . . . . .                       | 50         |
| 4.2      | Kernels . . . . .  | 55         |
| 4.3      | Which Loss Function? . . . . .                             | 57         |
| 4.4      | Confidence Sets . . . . .                                  | 57         |
| 4.5      | The Curse of Dimensionality . . . . .                      | 58         |
| 4.6      | Bibliographic Remarks . . . . .                            | 59         |
| 4.7      | Exercises . . . . .  | 59         |
| <b>5</b> | <b>Nonparametric Regression</b>                            | <b>61</b>  |
| 5.1      | Review of Linear and Logistic Regression . . . . .         | 62         |
| 5.2      | Linear Smoothers . . . . .                                 | 66         |
| 5.3      | Choosing the Smoothing Parameter . . . . .                 | 68         |
| 5.4      | Local Regression . . . . .                                 | 71         |
| 5.5      | Penalized Regression, Regularization and Splines . . . . . | 80         |
| 5.6      | Variance Estimation . . . . .                              | 85         |
| 5.7      | Confidence Bands . . . . .                                 | 88         |
| 5.8      | Average Coverage . . . . .                                 | 93         |
| 5.9      | Summary of Linear Smoothing . . . . .                      | 94         |
| 5.10     | Local Likelihood and Exponential Families . . . . .        | 95         |
| 5.11     | Scale-Space Smoothing . . . . .                            | 99         |
| 5.12     | Multiple Regression . . . . .                              | 100        |
| 5.13     | Other Issues . . . . .                                     | 111        |
| 5.14     | Bibliographic Remarks . . . . .                            | 119        |
| 5.15     | Appendix . . . . .   | 120        |
| 5.16     | Exercises . . . . .  | 121        |
| <b>6</b> | <b>Density Estimation</b>                                  | <b>125</b> |
| 6.1      | Cross-Validation . . . . .                                 | 126        |
| 6.2      | Histograms . . . . .                                       | 127        |
| 6.3      | Kernel Density Estimation . . . . .                        | 131        |
| 6.4      | Local Polynomials . . . . .                                | 137        |
| 6.5      | Multivariate Problems . . . . .                            | 138        |
| 6.6      | Converting Density Estimation Into Regression . . . . .    | 139        |
| 6.7      | Bibliographic Remarks . . . . .                            | 140        |
| 6.8      | Appendix . . . . .   | 140        |
| 6.9      | Exercises . . . . .  | 142        |
| <b>7</b> | <b>Normal Means and Minimax Theory</b>                     | <b>145</b> |
| 7.1      | The Normal Means Model . . . . .                           | 145        |
| 7.2      | Function Spaces . . . . .                                  | 147        |
| 7.3      | Connection to Regression and Density Estimation . . . . .  | 149        |

|           |   |            |
|-----------|---|------------|
| 7.4       | Stein's Unbiased Risk Estimator (SURE)                    | 150        |
| 7.5       | Minimax Risk and Pinsker's Theorem                        | 153        |
| 7.6       | Linear Shrinkage and the James-Stein Estimator            | 155        |
| 7.7       | Adaptive Estimation Over Sobolev Spaces                   | 158        |
| 7.8       | Confidence Sets   | 159        |
| 7.9       | Optimality of Confidence Sets                             | 167        |
| 7.10      | Random Radius Bands?                                      | 170        |
| 7.11      | Penalization, Oracles and Sparsity                        | 171        |
| 7.12      | Bibliographic Remarks                                     | 173        |
| 7.13      | Appendix  | 173        |
| 7.13.1    | The White Noise Model                                     | 173        |
| 7.13.2    | Weak Differentiability                                    | 174        |
| 7.13.3    | Proof of Pinsker's Theorem (Theorem 7.28)                 | 174        |
| 7.13.4    | Proof of Theorem 7.74                                     | 177        |
| 7.13.5    | Proof of Theorem 7.77                                     | 179        |
| 7.14      | Exercises   | 180        |
| <b>8</b>  | <b>Nonparametric Inference Using Orthogonal Functions</b> | <b>183</b> |
| 8.1       | Introduction  | 183        |
| 8.2       | Nonparametric Regression                                  | 183        |
| 8.3       | Irregular Designs   | 191        |
| 8.4       | Density Estimation  | 192        |
| 8.5       | Comparison of Methods                                     | 193        |
| 8.6       | Tensor Product Models                                     | 193        |
| 8.7       | Bibliographic Remarks                                     | 194        |
| 8.8       | Exercises   | 194        |
| <b>9</b>  | <b>Wavelets and Other Adaptive Methods</b>                | <b>197</b> |
| 9.1       | Haar Wavelets   | 199        |
| 9.2       | Constructing Wavelets                                     | 203        |
| 9.3       | Wavelet Regression  | 206        |
| 9.4       | Wavelet Thresholding                                      | 208        |
| 9.5       | Besov Spaces  | 213        |
| 9.6       | Confidence Sets   | 214        |
| 9.7       | Boundary Corrections and Unequally Spaced Data            | 215        |
| 9.8       | Overcomplete Dictionaries                                 | 216        |
| 9.9       | Other Adaptive Methods                                    | 216        |
| 9.10      | Do Adaptive Methods Work?                                 | 220        |
| 9.11      | Bibliographic Remarks                                     | 221        |
| 9.12      | Appendix  | 222        |
| 9.12.1    | Localization of Wavelets                                  | 222        |
| 9.12.2    | Fast Computations for Wavelets                            | 222        |
| 9.13      | Exercises   | 224        |
| <b>10</b> | <b>Other Topics</b>                                       | <b>227</b> |

|   |            |
|---|------------|
| 10.1 Measurement Error . . . . .          | 227        |
| 10.2 Inverse Problems . . . . .           | 233        |
| 10.3 Nonparametric Bayes . . . . .        | 235        |
| 10.4 Semiparametric Inference . . . . .   | 235        |
| 10.5 Correlated Errors . . . . .          | 236        |
| 10.6 Classification . . . . .             | 236        |
| 10.7 Sieves . . . . .                     | 237        |
| 10.8 Shape Restricted Inference . . . . . | 237        |
| 10.9 Testing . . . . .                    | 238        |
| 10.10 Computational Issues . . . . .      | 240        |
| 10.11 Exercises . . . . .                 | 240        |
| <b>Index</b>                              | <b>259</b> |