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 LaTeXmacros 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