

## Preface

The innovations in information and technology have revolutionized scientific research and knowledge discovery. They allow us to collect massive amount of data at relatively low cost. Observations with one-dimensional curves, two-dimensional images, and three-dimensional movies are commonly seen at the frontiers of scientific research. Statistical challenges arise from diverse fields of sciences, engineering, and humanities, ranging from genomic research, biomedical studies, and natural resource discovery to machine learning, financial engineering, and risk managements. High-dimensionality and massive amounts of data are the common features at the frontiers of statistical studies. These have profound impacts on statistical thinking, methodological development, and theoretical studies.

The 1980s and 1990s saw a phenomenal growth in numerous areas of statistical modeling and inferences. These areas include semiparametric models, data-analytic nonparametric methods, statistical learning, network tomography, analysis of longitudinal data, financial econometrics, time series, bootstrap and other resampling methodologies, statistical computing, mixed effects models, and robust multivariate analysis. This volume consists of 16 review articles and 6 research papers in these and some other related areas contributed by eminent leaders in statistics, who are the friends and former students of Professor Peter J. Bickel. It gives an overview of new developments and a contemporary outlook on the various frontiers of statistics. It is dedicated to Professor Peter John Bickel in honor of his 65th birthday. The monograph features an article by Kjell Doksum and Ya'acov Ritov that summarizes some of the Bickel's distinguished contributions. Most of the articles in the book will be presented at the "Workshop on Frontiers of Statistics", May 18 – 20, 2006, Princeton, New Jersey, co-sponsored by the Institute of Mathematical Statistics and International Indian Statistical Association, chaired by Jianqing Fan.

Ever since the publication of the monograph by Bickel, Klaassen, Ritov and Wellner (1993), the development in the asymptotically efficient

inference in semiparametric models has seen an exponential growth as is indicated in the chapter by Wellner, Klaassen and Ritov in this book. The contents of the chapter by Schick and Wefelmeyer review the developments of the asymptotically efficient inference in semiparametric time series models that have been developed only in the last 12 years and are on the cutting edge of this area of research. The chapter by Xia and Tong reviews numerous estimation methods in generalized linear regression models with an unknown link function, the so-called single index models, an area of research presently at the forefront. Van der Laan and Rubin discuss some innovative procedures based on cross validation for selecting estimating functions for estimation of parameters of interest. An oracle property is established under a general condition.

Closely related to the semiparametric modeling is the nonparametric methods for function estimation and their related applications. The area has been very active over the last twenty years, thanks to the exponential increase of computing power and availability of large data sets. Doksum and Schafer propose bandwidth selection procedures in local linear regression by maximizing the limiting power when testing the hypothesis of constant regression against local nonparametric Pitman alternatives. Among other things they show that the power optimal bandwidth can give much higher power than the bandwidth chosen by minimizing the mean squared error. The chapter by Hall and Kay has bearing on the classification of a given observation. They provide a new estimator of an index of atypicality and assess some of its optimality properties. Wang reviews wavelet statistical methodologies developed in the past fifteen years with emphasis on estimating regression functions, detecting and estimating change-points, solving statistical inverse problems, and studying self-similarity of a stochastic process. A classical problem in statistics is that of model diagnostics. In particular the most investigated model diagnostic problems in the literature are those of the lack-of-fit testing of a parametric regression model and the goodness-of-fit hypothesis of an error distribution in a given regression model. The chapter by Koul demonstrates that nonparametric techniques play an important role for lack-of-fit testing. It gives a brief review of asymptotically distribution free tests for these hypotheses based on certain marked empirical processes and residual empirical processes, respectively. The underlying theme is that all of these tests can be based on certain martingale transforms of these processes.

Related to nonparametric methods are boosting and bootstrap. An effective method of nonparametric classification and regression is boosting. The

paper of Bühlmann and Lutz provides a review of boosting and proposes a new bootstrap method for high-multivariate, linear time series. Lahiri reviews the literature on bootstrap methodology under independence and for different classes of dependent processes including Markov processes, long range dependent time series and spatial processes. An important tool for theoretical justification of bootstrap methods is asymptotic expansions. Such expansions are well understood when underlying distributions are either absolutely continuous or pure lattice distributions. Götze and van Zwet begin an investigation of discrete non-lattice distributions and discuss some new challenges in this interesting area of research.

The last two decades have seen a considerable amount of literature on the analysis of longitudinal and functional data. The chapter by Fan and Li presents an overview on recent developments in non- and semi-parametric inference for longitudinal data. Müller and Yao give a review of functional regression models and of the principal component analysis through a conditional expectation algorithm when having sparse longitudinal data.

Many exciting statistical developments over the last two decades lie at the interface between statistics and other subject disciplines. The chapter by Servidea and Meng demonstrates the fruitfulness of cross-fertilization between statistical physics and statistical computation, by focusing on the celebrated Swendsen-Wang algorithm for the Ising model and its recent perfect sampling implementation by Huber. Furthermore, it outlines some important results and open problems in these areas. Lawrence, Michailidis, Nair and Xi provide a review of the statistical issues and developments in network tomography with an emphasis on active tomography and illustrate the results with an application to internet telephony. A comprehensive overview on likelihood inferences for diffusion processes is given by Aït-Sahalia, based on discretely sampled data. This is an important subject in financial econometrics where statistics plays an important role. Aït-Sahalia gives extensive treatments on the expansions of transition densities and their applications. The chapter by Park and Jeong gives a review of the recently developed theory for several promising nonparametric estimators of frontiers or boundaries in productivity analysis.

Many contemporary statistical problems borrow the ideas from traditional statistics. Ghosh and Tokdar presents some of the recent development on the convergence and consistency proof of Newton's algorithm for estimating a mixing distribution. Jiang and Ge give an overview of linear, generalized linear and nonlinear mixed effects models with emphases on recent developments and challenges. It is well known that the sample mean

vector and the sample covariance matrix are optimal when the underlying data are normal but extremely sensitive to outliers and heavy tailed data. The chapter by Zuo surveys numerous robust alternatives of these classical location and scatter estimators and discusses their applications to the multivariate data analysis. Wong proposes an innovative idea for estimating the loss of an estimate.

All the chapters in the book have been reviewed by world experts. We take this opportunity to thank all the referees for rendering their invaluable help: Sarat Dass, Irene Gijbels, Jayanta Ghosh, Heng Peng, Marian Hristache, Jinhua Huang, Ildar Ibragimov, Gareth James, Jiming Jiang, Soumendra Lahiri, Linyuan Li, Yin Li, Shiqing Ling, Neal Madras, Moushen Pourhamadi, Anton Schick and Chunming Zhang. We are most grateful to the enthusiastic support of all the people who have helped us. We are particularly indebted to Dr. Heng Peng and Dr. Chongqi Zhang for invaluable help and editorial assistance, turning individual contributions into such a wonderful book. Assistance from graduate students at Princeton, including Yingying Fan, Clifford Lam, Yue Niu and Jinchi Lv are gratefully acknowledged. The generous financial supports from National Science Foundation, Minerva Foundation, Bendheim Center for Finance, and Department of Operations Research and Financial Engineering make the workshop and this volume possible.

Professor Peter J. Bickel has made numerous significant contributions to many of the fields discussed in this volume as is evidenced by references to his work. We present this volume to him in celebration of his 65th birthday. We look forward to his many more innovative contributions to statistical science.

Jianqing Fan, Princeton  
Hira L. Koul, East Lansing  
December 22, 2005