PREFACE TO THE 2015 EDITION

In this preface, we start with an overview of developments in statistics since the first (1977) edition, then give separate overviews of Volumes I and II of the second edition.

In the last 40 some years statistics has changed enormously under the impact of several forces:

(1) The generation of what were once unusual types of data such as images, trees (phylogenetic and other), and other types of combinatorial objects.

(2) The generation of enormous amounts of data—terabytes (the equivalent of 10^{12} characters) for an astronomical survey over three years.

(3) The possibility of implementing computations of a magnitude that would have once been unthinkable.

The underlying sources of these changes have been the exponential change in computing speed (Moore's "law") and the development of devices (computer controlled) using novel instruments and scientific techniques (e.g., NMR tomography, gene sequencing). These techniques often have a strong intrinsic computational component. Tomographic data are the result of mathematically based processing. Sequencing is done by applying computational algorithms to raw gel electrophoresis data.

As a consequence the emphasis of statistical theory has shifted away from small sample optimality results in a number of directions:

(1) Methods for inference based on larger numbers of observations and minimal assumptions—asymptotic methods in non- and semiparametric models, models with "infinite" number of parameters.

(2) The construction of models for time series, temporal spatial series, and other complex data structures using sophisticated probability modeling but again relying for analytical results on asymptotic approximation. Multiparameter models are the rule.

(3) The use of methods of inference involving simulation as a key element such as the bootstrap and Markov Chain Monte Carlo.

(4) The development of techniques not describable in "closed mathematical form" but rather through elaborate algorithms for which problems of existence of solutions are important and far from obvious.

(5) The study of the interplay between numerical and statistical considerations. Despite advances in computing speed, some methods run quickly in real time. Others do not and some though theoretically attractive cannot be implemented in a human lifetime.

(6) The study of the interplay between the number of observations and the number of parameters of a model and the beginnings of appropriate asymptotic theories.

There have been other important consequences such as the extensive development of graphical and other exploratory methods for which theoretical development and connection with mathematics have been minimal. These will not be dealt with in our work.

In this edition we pursue our philosophy of describing the basic concepts of mathematical statistics relating theory to practice.

Volume I

This volume presents the basic classical statistical concepts at the Ph.D. level without requiring measure theory. It gives careful proofs of the major results and indicates how the theory sheds light on the properties of practical methods. The topics include estimation, prediction, testing, confidence sets, Bayesian analysis and the more general approach of decision theory.

We include from the start in Chapter 1 non- and semiparametric models, then go to parameters and parametric models stressing the role of identifiability. From the beginning we stress function-valued parameters, such as the density, and function-valued statistics, such as the empirical distribution function. We also, from the start, include examples that are important in applications, such as regression experiments. There is extensive material on Bayesian models and analysis and extended discussion of prediction and k-parameter exponential families. These objects that are the building blocks of most modern models require concepts involving moments of random vectors and convexity that are given in Appendix B.

Chapter 2 deals with estimation and includes a detailed treatment of maximum likelihood estimates (MLEs), including a complete study of MLEs in canonical k-parameter exponential families. Other novel features of this chapter include a detailed analysis, including proofs of convergence, of a standard but slow algorithm (coordinate descent) for convex optimization, applied, in particular to computing MLEs in multiparameter exponential families. We also give an introduction to the EM algorithm, one of the main ingredients of most modern algorithms for inference. Chapters 3 and 4 are on the theory of testing and confidence regions, including some optimality theory for estimation as well and elementary robustness considerations.

Chapter 5 is devoted to basic asymptotic approximations with one dimensional parameter models as examples. It includes proofs of consistency and asymptotic normality and optimality of maximum likelihood procedures in inference and a section relating Bayesian and frequentist inference via the Bernstein–von Mises theorem.

Finally, Chapter 6 is devoted to inference in multivariate (multiparameter) models. Included are asymptotic normality and optimality of maximum likelihood estimates, inference in the general linear model, Wilks theorem on the asymptotic distribution of the likelihood ratio test, the Wald and Rao statistics and associated confidence regions, and some parallels to the optimality theory and comparisons of Bayes and frequentist procedures given in the one dimensional parameter case in Chapter 5. Chapter 6 also develops the asymptotic joint normality of estimates that are solutions to estimating equations and presents Huber's Sandwich formula for the asymptotic covariance matrix of such estimates. Generalized linear models, including binary logistic regression, are introduced as examples. Robustness from an asymptotic theory point of view appears also. This chapter uses multivariate calculus in an intrinsic way and can be viewed as an essential prerequisite for the more advanced topics of Volume II.

Volume I includes Appendix A on basic probability and a larger Appendix B, which includes more advanced topics from probability theory such as the multivariate Gaussian distribution, weak convergence in Euclidean spaces, and probability inequalities as well as more advanced topics in matrix theory and analysis. The latter include the principal axis and spectral theorems for Euclidean space and the elementary theory of convex functions on R^d as well as an elementary introduction to Hilbert space theory. As in the first edition, we do not require measure theory but assume from the start that our models are what we call "regular." That is, we assume either a discrete probability whose support does not depend on the parameter set, or the absolutely continuous case with a density. Hilbert space theory is not needed, but for those who know this topic Appendix B points out interesting connections to prediction and linear regression analysis.

Appendix B is as self-contained as possible with proofs of most statements, problems, and references to the literature for proofs of the deepest results such as the spectral theorem. The reason for these additions are the changes in subject matter necessitated by the current areas of importance in the field.

For the first volume of the second edition we would like to add thanks to Jianging Fan, Michael Jordan, Jianhua Huang, Ying Qing Chen, and Carl Spruill and the many students who were guinea pigs in the basic theory course at Berkeley. We also thank Faye Yeager for typing, Michael Ostland and Simon Cawley for producing the graphs, Yoram Gat for proofreading that found not only typos but serious errors, and Prentice Hall for generous production support.

Volume II

Volume II of the second edition will be forthcoming in 2015. It presents what we think are some of the most important statistical concepts, methods, and tools developed since the first edition. Topics to be included are: asymptotic efficiency in semiparametric models, semiparametric maximum likelihood estimation, survival analysis including Cox regression, classification, methods of inference based on sieve models, model selection, Monte Carlo methods such as the bootstrap and Markov Chain Monte Carlo, nonparametric curve estimation, and machine learning including support vector machines and classification and regression trees (CART).

The basic asymptotic tools that will be developed or presented, in part in the text and, in part in appendices, are weak convergence for random processes, elementary empirical process theory, and the functional delta method.

With the tools and concepts developed in this second volume students will be ready for advanced research in modern statistics.

We thank Akichika Ozeki and Sören Künzel for pointing out errors and John Kimmel and CRC Press for production support. We also thank Dee Frana and especially Anne Chong who typed 90% of Volume II for word processing.

Last and most important we would like to thank our wives, Nancy Kramer Bickel and Joan H. Fujimura, and our families for support, encouragement, and active participation in an enterprise that at times seemed endless, appeared gratifyingly ended in 1976 but has, with the field, taken on a new life.

Peter J. Bickel
bickel@stat.berkeley.edu
Kjell Doksum
doksum@stat.wisc.edu

CONTENTS

VOLUME II CONTENTS

This page intentionally left blank