WEAK CONVERGENCE
AND
EMPIRICAL PROCESSES
With Applications to Statistics

by Aad van der Vaart and Jon A. Wellner

January 1996

Preface

This book tries to do three things. The first goal is to give an exposition of certain modes of stochastic convergence, in particular convergence in distribution. The classical theory of this subject was developed mostly in the 1950s and is well summarized in Billingsley (1968). During the last 15 years, the need for a more general theory allowing random elements that are not Borel measurable has become well established, particularly in developing the theory of empirical processes. Part 1 of the book, *Stochastic Convergence*, gives an exposition of such a theory following the ideas of J. Hoffmann-Jørgensen and R. M. Dudley.

A second goal is to use the weak convergence theory background developed in Part 1 to present an account of major components of the modern theory of empirical processes indexed by classes of sets and functions. The weak convergence theory developed in Part 1 is important for this, simply because the empirical processes studied in Part 2, *Empirical Processes*, are naturally viewed as taking values in nonseparable Banach spaces, even in the most elementary cases, and are typically *not* Borel measurable. Much of the theory presented in Part 2 has previously been scattered in the journal literature and has, as a result, been accessible only to a relatively small number of specialists. In view of the importance of this theory for statistics, we hope that the presentation given here will make this theory more accessible to statisticians as well as to probabilists interested in statistical applications.

Our third goal is to illustrate the usefulness of modern weak convergence theory and modern empirical process theory for statistics by a wide variety of applications. On the one hand, as is also clear through the work of David Pollard, the theory of empirical processes provides a collection of extremely powerful tools for proving many of the main limit theorems of asymptotic statistics. On the other hand, the empirical distribution indexed by a collection of sets or functions is an object of independent statistical interest; for instance, as a measure of goodness of fit. The topics included in Part 3 of the book, *Statistical Applications*, range from rates of convergence in semiparametric estimation, to the functional delta-method, bootstrap, permutation empirical processes, and the convolution theorem. We have not aimed at giving an exhaustive coverage of the statistical background or related literature in presenting these applications. The choices made reflect our personal interests and research efforts over the past few years, so the reader should understand that many equally interesting applications and further research directions are not covered here. For instance, we expect significant progress in semiparametric theory through the use of empirical processes in the next few years. Wellner (1992) reviews various applications of empirical process methods through 1992.

This project began with a joint effort to develop some results in the Hoffmann-Jørgensen theory (namely Prohorov's Theorem 1.3.9) needed to prove the convolution and asymptotic minimax theorem for estimators with values in nonseparable Banach spaces (see Chapter 3.11). That effort was successful, but it resulted in a manuscript that was too awkward for publication in a journal. The generality offered by the new weak convergence theory and the many applications of general empirical process theory in statistics led us to explore the area further. The result, several years later, is this book.

Along the way we have learned much from the main contributors to empirical process theory, from colleagues, and from friends. In particular, we owe thanks to Lucien Birgé, R. M. Dudley, Peter Gaenssler, Sara van de Geer, Richard Gill, Evarist Giné, Piet Groeneboom, Lucien Le Cam, Michel Ledoux, Pascal Massart, Susan Murphy, David Pollard, Michel Talagrand, and Joel Zinn, for advice, corrections, discussions, clarifications, inspiration, preprints, and help.

The authors presented preliminary versions of various parts of the manuscript in courses at the Free University, Amsterdam, and the University of Washington to patient and sharp-eyed groups of students, including Jian Huang, Marianne Jonker, Brad McNeney, Jinko Graham, Jens Praestgaard, and Anne Sheehy. We owe them as well as Franz Strobl, Klaus Ziegler, and Lutz Duembgen thanks for picking up errors and points needing clarification.

The authors have been partially supported in their research efforts over the period of work on this book (1988–1995) by grants from the National Science Foundation, NIAID, NWO, NATO, CNRS, Stieltjes Institute, and Free University Amsterdam (which funded several visits of the second author to that fair city). Our thanks also go to our editor at Springer, Martin

Gilchrist, and to the staff at Springer-Verlag for their efficient job in turning the TEX manuscript into a published book.

Amsterdam and Seattle
August 1995

Reading Guide

This book consists of three parts and an appendix. Part 1 is an exposition of (mostly) three modes of convergence of stochastic variables with values in metric spaces: convergence in distribution, convergence in probability, and almost sure convergence. A new aspect of this part, as compared to existing literature, is the fact that the stochastic variables are not assumed measurable with respect to the Borel σ-field. Part 1 is also useful in collecting a large number of results that have thus far not been available in book form in sufficient generality. In particular this concerns the systematic treatment of convergence in distribution in spaces of bounded functions equipped with the supremum metric. Most subjects treated in Part 1 are used in the two later parts of the book, but Chapters 1.2, 1.3, 1.5, 1.9, and 1.10 probably form the core of this part, and should be read in this order. Part 1 begins with a thorough introduction.

Part 2 is mostly concerned with the empirical measure and empirical process of a sample of observations, indexed by a class of functions. These are the maps $f \mapsto \sum_{i=1}^n f(X_i)$ and $f \mapsto n^{-1/2}\sum_{i=1}^n \bigl(f(X_i) - Pf\bigr)$ whose domain is a class $\mathcal{F}$ of measurable functions. The main results in this part are contained in Chapters 2.4 and 2.5 and concern the uniform law of large numbers (Glivenko-Cantelli theorem) and uniform central limit theorem (Donsker theorem), respectively. Chapters 2.6, 2.7, and 2.10 contain many examples of classes that satisfy the conditions of the theorems in these chapters. Chapter 2.1 is an introduction and Chapters 2.2 and 2.3 are necessary preparation for the main results of Part 2. Again, many of the results from the other chapters reappear later in the book but need not be studied in sequential order.

Part 3 consists of 11 chapters, which in principle can be read independently. This part shows the wide range of applications of the results obtained in the earlier parts in statistics, ranging from parametric and nonparametric estimation to the bootstrap, the functional delta-method, and Kolmogorov-Smirnov statistics. Every chapter assumes familiarity of the basic notation of Parts 1 and 2, but no section requires knowledge of more than a few sections of Part 2. Chapter 3.1 gives an overview of this part.

The material presented in the three parts is self-contained to a reasonable extent. The appendix covers a number of auxiliary subjects that are used to develop some of the material in the three main Parts. Some results are presented with proof and some without, to serve as an easy reference.

Most of the chapters contain a number of "problems and complements" at the end. A number of these are real, textbook-style problems, but a lot of this material is meant as a supplement to the main text. Some problems present technical details, while other problems concern additional results of interest. We should warn the reader that the density of errors in these problem sections is probably higher than in the main text. Many problems and complements have not been double-checked.

Contents