*NSF-CBMS Regional Conference Series*
*in Probability and Statistics*
*Volume 2*

# EMPIRICAL PROCESSES: THEORY AND APPLICATIONS

**David Pollard**
*Yale University*

Institute of Mathematical Statistics
Hayward, California

American Statistical Association
Alexandria, Virginia

Conference Board of the Mathematical Sciences

*Regional Conference Series*
*in Probability and Statistics*

Supported by the
National Science Foundation

The production of the *NSF-CBMS Regional Conference Series in Probability and Statistics* is managed by the Institute of Mathematical Statistics: Paul Shaman, IMS Managing Editor; Jessica Utts, IMS Treasurer; and Jose L. Gonzalez, IMS Business Manager.

Library of Congress Catalog Card Number: 90-50461

International Standard Book Number 0-940600-16-1



Printed in the United States of America

To Gai and James

# Contents

# Preface

These notes grew from lectures I gave at the University of Iowa in July of 1988, as part of the *NSF-CBMS Regional Conference Series*. The conference was ably organized by Tim Robertson and Richard Dykstra. I am most grateful to them for giving me the opportunity to experiment on a live and receptive audience with material not entirely polished. I also appreciate the suggestions and comments of Richard Dudley. Much of the lecture material was repackaging of ideas originally due to him.

In reworking the lecture notes I have tried (not always successfully) to resist the urge to push the presentation to ever higher levels of generality. My aim has been to introduce just enough technique to handle typical nontrivial asymptotic problems in statistics and econometrics. Of course the four substantial examples that represent the applications part of the lectures do not exhaust the possible uses for the theory. I have chosen them because they cleanly illustrate specific aspects of the theory, and also because I admire the original papers.

To anyone who is acquainted with the empirical process literature these notes might appear misleadingly titled. Empirical process theory usually deals with sums of independent (identically distributed) random variables $f(\xi_i(\omega))$, with $f$ running over a class of functions $\mathcal{F}$. However I have chosen to present results for sums of independent stochastic processes $f_i(\omega, t)$ indexed by a set $T$. Such a setting accommodates not only the relatively straightforward generalization to nonidentically distributed $\{\xi_i\}$, but also such simple modifications as a rescaling of the summands by a factor that depends on $i$ and $\omega$. It has often irked me that the traditional notation cannot handle summands such as $f(\xi_i)/i$, even though the basic probabilistic method is unaffected.

The cost of the modified notation appears in two ways. Some familiar looking objects no longer have their usual meanings. For example, $\mathcal{F}$ will now stand for a subset of $\mathbb{R}^n$ rather than for a class of functions. Also, some results, such as the analogues in Section 4 of the standard Vapnik-Červonenkis theory, become a trifle less general than in the traditional setting. The benefits include the natural reinterpretation of the Vapnik-Červonenkis property as a sort of dimensionality concept, and the transformation of $\mathcal{L}^2(P_n)$ pseudometrics on classes of functions into the usual ($\ell_2$) Euclidean distances in $\mathbb{R}^n$.

Several friends and colleagues at Yale and elsewhere have influenced the final form of the notes. Ariel Pakes provided well thought-out comments on the paper Pollard (1989), in which I tried out some of the ideas for the Iowa lectures. Probing questions from Don Andrews firmed up some particularly flabby parts of the original lecture notes. A faithful reading group struggled through the first half of the material, finding numerous errors in what I had thought were watertight arguments. Deborah Nolan tested a slightly more correct version of the notes on a graduate class at the University of California, Berkeley. (The rate at which bugs appeared suggests there might even be other embarrassing errors lying in wait to confuse future unsuspecting readers.) I thank them all.

Very recently I had the good fortune to obtain a copy of the manuscript by Ledoux and Talagrand (1990), which provides an alternative (often mathematically more elegant) treatment for some of the material in the theory part of the notes. I am grateful to those authors for enlightening me.

As always Barbara Amato well deserves my thanks and admiration for her ability to convert unruly drafts into beautiful TEXnical documents. Paul Shaman and Jose Gonzalez contributed valuable editorial advice. The manuscript was prepared using the AMS-TEX macros of the American Mathematical Society.